ANALYSIS OF 13 *TP53* AND *WRAP53* POLYMORPHISM FREQUENCIES IN RUSSIAN POPULATIONS

Olkova MV[1,2] ✉, Petrushenko VS[2], Ponomarev GYu[2]

[1] Research Centre of Medical Genetics (RCMG), Moscow, Russia

[2] Vavilov Institute of General Genetics, Moscow, Russia

In the last decade the search for and annotation of human genome polymorphisms associated with phenotype have become particularly important concerning the opportunity of their use in medical and population genetics, pharmacogenomics and evolutionary biology. The study was aimed to calculate the frequencies and analyze the prevalence of 13 germline polymorphisms of two genes, *TP53* encoding the genome-keeper p53 protein and *WRAP53* involved in regulation of p53 production, in 28 Russian populations. We obtained data on 9 exonic *TP53* variants (rs587781663, rs17882252, rs150293825, rs112431538, rs149633775, rs144340710, rs1042522, rs1800371, rs201753350), one intronic polymorphism (rs17881850), and three variants of *WRAP53* (rs17880282, rs2287499, rs34067256). In the majority of populations the sample size was over 50 people (except five populations with 30–49 surveyed people). The alternative alleles' population frequencies for studies genetic variants in most Russian populations were close to appropriate allele frequencies in European and Asian populations of similar origin taken from global databases. The exceptions were six populations ("Central Caucasus", "Dagestan", "northern Russians", "southeastern Russians", "Tatars" and "Transcaucasia") with increased alternative alleles' population frequencies. All listed populations except the population of "southeastern Russians" are characterized by polymorphisms with high allele frequencies not satisfying the Hardy–Weinberg principle.

Keywords: p53, TP53, WRAP53, tumor marker, polymorphism, population frequency, genetic epidemiology

Acknowledgement: we would like to express our appreciation to Oleg Balanovsky, head of the Genome Geography Laboratory of the Vavilov Institute of General Genetics for study management and manuscript editing, to all DNA donors and Biobank of North Eurasia for provided collection of samples, as well as to the Center for Precision Genome Editing and Genetic Technologies for Biomedicine of the Pirogov Russian National Research Medical University (Moscow, Russia) for the opportunity to use the molecular biology technologies.

Funding: the study was carried out as part of the public contract between the Ministry of Science and Higher Education of the Russian Federation and the Research Centre of Medical Genetics (phenotyping of samples, database construction, data analysis).

Author contribution: Olkova MV — study design, statistical analysis, manuscript writing; Petrushenko VS — bioinformatics analysis, Ponomarev GYu — experiments.

Compliance with ethical standards: the study was carried out in accordance with the World Medical Association Declaration of Helsinki. All samples were obtained from Biobank of North Eurasia. The informed consent was obtained from all donors.

✉ **Correspondence should be addressed:** Marina V. Olkova
Gubkina 3, Moscow, 119991; genetics@inbox.ru

Received: 26.11.2020 **Accepted:** 12.12.2020 **Published online:** 12.01.2021

DOI: 10.24075/brsmu.2021.001

АНАЛИЗ ЧАСТОТ 13 ПОЛИМОРФИЗМОВ В ГЕНАХ *TP53* И *WRAP53* В РОССИЙСКИХ ПОПУЛЯЦИЯХ

М. В. Олькова[1,2] ✉, В. С. Петрушенко[2], Г. Ю. Пономарев[2]

[1] Медико-генетический научный центр имени академика Н. П. Бочкова, Москва, Россия

[2] Институт общей генетики имени Н. И. Вавилова РАН, Москва, Россия

В последнее десятилетие поиск и аннотация ассоциированных с фенотипом геномных полиморфизмов человека, а также изучение их популяционных частот стали особенно актуальными в связи с возможностью их применения в медицинской и популяционной генетике, фармакогеномике и эволюционной биологии. Целью исследования было рассчитать частоту и проанализировать распространенность в 28 российских популяциях 13 герминальных полиморфизмов двух генов — *TP53,* матрицы «хранителя генома» белка p53 и гена *WRAP53*, влияющего на производство белка p53. Были получены данные для 9 экзонных вариантов гена *TP53* (rs587781663, rs17882252, rs150293825, rs112431538, rs149633775, rs144340710, rs1042522, rs1800371, rs201753350), одного интронного полиморфизма (rs17881850), а также трех вариантов гена *WRAP53* (rs17880282, rs2287499, rs34067256). Для большинства популяций выборка была представлена числом более 50 человек (за исключением пяти популяций, в которых было обследовано от 30 до 49 человек). Популяционные частоты альтернативных аллелей изученных генных вариантов в большинстве российских популяций оказались близки к значениям частот этих аллелей в соответствующей их происхождению европейской или азиатской популяции из мировых баз данных. Исключение составили шесть популяций («Центральный Кавказ», «Дагестан», «северные русские», «юго-восточные русские», «татары» и «Закавказье»), в которых популяционные частоты альтернативных аллелей для большинства маркеров оказались повышенными. Для всех вышеперечисленных популяций, кроме «юго-восточных русских», характерно несоответствие аллелей полиморфизмов с повышенными частотами равновесию Харди–Вайнберга.

Ключевые слова: p53, TP53, WRAP53, онкомаркер, полиморфизм, популяционная частота, генетическая эпидемиология

Благодарности: О. П. Балановскому, заведующему лабораторией геномной географии Института общей генетики им. Н. И. Вавилова, за руководство исследованием и правку статьи, всем донорам ДНК и АНО «Биобанк Северной Евразии» за предоставленную коллекцию образцов, а также Центру высокоточного редактирования и генетических технологий для биомедицины РНИМУ им. Н. И. Пирогова (Москва, Россия) за возможность использования молекулярно-генетических технологий.

Финансирование: исследование выполнено в рамках Государственного задания Министерства науки и высшего образования РФ для Медико-генетического научного центра им. академика Н. П. Бочкова (работы по фенотипированию образцов, созданию базы данных, анализу данных).

Вклад авторов: М. В. Олькова — дизайн, статистический анализ, написание текста статьи; В. С. Петрушенко — биоинформатический анализ, Г. Ю. Пономарев — экспериментальные работы.

Соблюдение этических стандартов: исследование проведено в соответствии с требованиями Хельсинкской декларации Всемирной медицинской ассоциации. Все образцы для исследования получены из «Биобанка Северной Евразии». От всех доноров получено добровольное информированное согласие.

✉ **Для корреспонденции:** Марина Викторовна Олькова
ул. Губкина, д. 3, г. Москва, 119991; genetics@inbox.ru

Статья получена: 26.11.2020 **Статья принята к печати:** 12.12.2020 **Опубликована онлайн:** 12.01.2021

DOI: 10.24075/vrgmu.2021.001

TP53 gene is responsible for synthesis of one of the most notorious tumor suppressors, the p53 protein, which plays a vital part in maintaining genetic stability of the cell and cancer prevention. After activation due to cell damage, p53 triggers a number of cellular responses aimed at cell recovery and survival, or, in case the recovery is impossible, at programmed cell death. Such diverse pleiotropic tissue effects of p53 are due to total effect of co-expressed p53 isoforms. To date, at least 12 p53 isoforms have been reported, which are produced through alternative initiation of translation, the alternative promoter usage and alternative splicing [1]. All p53 isoforms share the common DNA-binding domain, but contain distinct transactivation and inhibitory domain, enabling the differential regulation of gene expression [2].

The *TP53* gene shows an autosomal dominant pattern of inheritance; it is associated with the risk of Li–Fraumeni syndrome and other hereditary cancer syndromes. The altered sensitivity to certain medications in people with a number of TP53 gene polymorphisms has been confirmed (Table 1).

The region of *WRAP53* gene with at least three alternative promoters is located in the region 13.1 of the short arm of chromosome 17 partially overlapping the 5'-region of the TP53 gene, which is located on the chain oppositely oriented to *WRAP53* in a head-to-head orientation [3]. The *WRAP53* gene plays a dual role. First, it encodes the antisense RNA (WRAP53α), which regulates the levels of p53 mRNA through interaction with the first exon of *TP53*, and is also involved in stimulation of p53 protein production due to its impact on the 5'- untranslated region of p53 mRNA [4, 5]. Second, *WRAP53* is responsible for WRAP53β protein (also called WDR79 and TCAB1) synthesis; WRAP53β belongs to WD40 protein family. This protein contributes to maintaining the integrity and normal function of the Cajal bodies essential for maturation of the splicing machinery and telomere maintenance [6–8]. WRAP53β also promotes accumulation of the repair factor 53BP1 at DNA double-strand breaks, thus stimulating the DNA repair [9]. The WRAP53β protein possibly possesses oncogenic properties, as evidenced by the WRAP53β overexpression in various cancer cell lines compared to normal cells [7, 8]. It should be noted that the involvement of this protein in carcinogenesis currently remains questionable: there is a theory that overexpression may be caused by the involvement of WRAP53β in DNA multiple double-strand break repair in case of cancer development in certain tissue [7].

The *WRAP53* gene mutations show the autosomal recessive pattern of inheritance. The homozygous mutations of this gene may result in dyskeratosis congenita and Li–Fraumeni syndrome.

The clinical significance of genes *TP53* and *WRAP53*, as well as high prevalence of their germline pathogenic variants in various cancer types [10], explain the need for studying the frequencies of these genes in populations of different countries. The frequencies of polymorphisms of these genes have already been studied in some European countries and the United States: the details on frequencies of both clinically significant polimorphisms and markers with uncertain significance may be found on the web-sites of such projects as ClinVar [11] of the National Center for Biotechnology Information of the USA, Ensembl (joint scientific project of the European Bioinformatics Institute and Sanger Institute) [12], and Genome Aggregation Database (gnomAD) [13]. In Russia, "Genokarta", the web-site of genetic encyclopedia created by researchers from the Novosibirsk State University is being actively developed [14]. Our study aimed at exploring distribution and frequencies of 13 polymorphisms of *TP53* and *WRAP53* genes in Russian populations is directed to expand the scientific knowledge in this area with regard to populations living in our country.

METHODS

DNA sampling

DNA samples were provided by Biobank of North Eurasia [15]. DNA was extracted from blood and saliva using the standard phenol-chloroform extraction method. The study included 1,785 DNA samples of volunteers who belonged to 28 Russian populations, which, based on their places of residence, covered the main regions of Russia (Table 2). Inclusion criteria: volunteers belonging to certain ethnic group (based on the self-identified ethnicity in four or more generations). Exclusion criteria: samples failing to meet criteria of belonging to certain ethnic group. Since the study was aimed at investigation of autosomal markers, the gender distribution was not taken into account during DNA sampling.

The size of each population was 30–87 people. The composition of studied populations is presented in Table 2. It should be taken into account that the studied genes *TP53* and *WRAP53* were autosomal, therefore, the actual number of studied alleles was twice as much: 60–174 alleles for each population.

Selection of polymorphisms

The list of *TP53* and *WRAP53* polymorphisms was established based on genetic variants with proven clinical significance submitted to ClinVar database (except the TP53 intronic variant, rs17881850). The intronic variant rs17881850 was included in the study in order to compare the allele frequencies of neutral polymorphism with the allele frequencies of genetic variants with confirmed clinical significance. Unfortunately, after genotyping much of polymorphisms included in original list had to be excluded from analysis, i. e. the population frequencies were calculated only using markers genotyped successfully in all populations.

Genotyping

All individuals were genotyped for nine *TP53* exon polymorphisms (rs587781663, rs17882252, rs150293825, rs112431538, rs149633775, rs144340710, rs1042522, rs1800371, rs201753350) and one intronic variant (rs17881850), as well as for three *WRAP53* polymorphisms (rs17880282, rs2287499, rs34067256). Genotyping was carried out using the Illumina (Illumina Inc.; USA) genomic analysis microarray technology. The standard 0.15 GenCall score cutoff value was used to discard the poorly typed samples.

Basic data on studied polymorphisms

Full information on studied polymorphisms was obtained from the web-site of the National Center for Biotechnology Information of the USA [16], in particular from the ClinVar archive [17] and Genome Aggregation Database (gnomAD) [18]. The position of polymorphism in the human genome was specified based on the version GRCh38.p12 of human reference Genome Assembly (Table 1).

Since the information about some polymorphisms available from public domains was incomplete, all markers were also studied by functional analysis through hidden Markov models designed for prediction of missense protein variants using the

Table 1. Basic data on studied markers; data obtained from ClinVar (NCBI)

Gene	Marker	GRCh38.p12	Gene region	Reference nucleotide	Alternative nucleotide	Polymorphism type	Amino acid substitution	Clinical significance of polymorphism (ClinVar)	FATHMM analysis results	Marker validation	Predisposition to disease
TP53	rs17881850	chr17:7669739	Intron 10	G	A	Intronic	No	No information	No calculation	No	No
	rs587781663	chr17: 7670627	Exon 10	C	T	Missense	Gly361Glu	No information	No calculation	Validated	Li–Fraumeni syndrome and other hereditary cancer syndromes
	rs17882252	chr17: 7670694	Exon 10	C	T	Missense	Glu339Lys	Likely benign/uncertain significance	CANCER	Validated. Conflicting data	
	rs150293825	chr17:7670695	Exon 10	G	A	Silent	Phe338=	Benign/likely benign	CANCER	Validated. No conflicting data	
	rs112431538	chr17:7673767	Exon 10	C	T	Missense	Glu285Lys	Pathogenic/ likely pathogenic	CANCER	Approved by FDA experts	
	rs149633775	chr17:7673773	Exon 7	G	A	Missense	Arg283Cys	Uncertain significance	CANCER	Проверен экспертами FDA	
	rs144340710	chr17:7674259	Exon 6	T	C	Missense	Asn235Ser	Benign	CANCER	Approved by FDA experts	
	rs1042522	chr17:7676154	Exon 3	G	C	Missense	Pro72Arg	Benign	CANCER	Included in expert panel of Pharmacogenomics Knowledgebase (PharmGKB)	
	rs1800371	chr17:7676230	Exon 3	G	A	Missense	Pro47Ser	Benign	CANCER	Approved by FDA experts	
	rs201753350	chr17:7676387	Exon 2	C	T	Missense	Val31Ile	Benign/likely benign/ conflicting data	CANCER	Validated. Conflicting data	
WRAP53	rs17880282	chr17:7688679	Exon 2	C	T	Missense	Pro11Ser	Benign/likely benign	PASSENGER/ OTHER	Validated. No conflicting data	Li–Fraumeni syndrome, dyskeratosis congenita
	rs2287499	chr17:7688850	Exon 2	C	G	Missense	Arg68Gly	Benign	PASSENGER/ OTHER	Validated. No conflicting data	
	rs34067256	chr17:7689055	Exon 2	C	G	Missense	Pro136Arg	Benign/likely benign	PASSENGER/ OTHER	Validated. No conflicting data	

Table 2. *TP53* and *WRAP53* polymorphism frequencies in studied Russian populations and reference worlds' populations

Population	Численность	rs17881850	rs587781663	rs17882252	rs150293825	rs112431538	rs149633775	rs144340710	rs1042522	rs1800371	rs201753350	rs17880282	rs2287499	rs34067256
Altaians	77	0.026	0.000	0.007	0.000	0.000	0.000	0.000	0.766	0.000	0.000	0.006	0.156	0.000
Bashkirs	43	0.000	0.012	0.000	0.000	0.000	0.000	0.000	0.720	0.000	0.000	0.000	0.116	0.000
Belarusians and Russians of the Northwest	30	0.017	0.000	0.000	0.000	0.000	0.000	0.017	0.650	0.000	0.000	0.000	0.117	0.033
Buryats, Khamnigans, Yakuts	57	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.772	0.000	0.000	0.000	0.158	0.000
Central Caucasus	64	0.071	0.048	0.063	0.047	0.032	0.032	0.000	0.697	0.032	0.032	0.047	0.125	0.016
Chukchis and Koryaks	67	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.657	0.000	0.000	0.000	0.313	0.000
Dagestan	79	0.025	0.013	0.019	0.013	0.013	0.013	0.000	0.643	0.019	0.025	0.013	0.165	0.025
Far East (peoples of the Amur region)	84	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.690	0.000	0.000	0.012	0.208	0.000
Karelians and Vepsians	59	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.717	0.000	0.000	0.000	0.113	0.000
Kazakhs, Karakalpaks, Uighurs, Nogays	44	0.025	0.000	0.008	0.000	0.000	0.000	0.000	0.797	0.000	0.000	0.000	0.042	0.000
Khanty, Mansi, Nenets	53	0.034	0.011	0.000	0.000	0.000	0.000	0.000	0.727	0.011	0.000	0.000	0.216	0.000
Komi and Udmurts	84	0.006	0.006	0.006	0.000	0.012	0.000	0.000	0.645	0.006	0.000	0.006	0.232	0.000
Mari and Chuvash	53	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.651	0.000	0.000	0.000	0.264	0.009
Khalkha Mongols	49	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.688	0.000	0.000	0.000	0.198	0.020
Mongols (other groups) and Kalmyks	78	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.654	0.000	0.000	0.000	0.231	0.000
Mordvins	40	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.700	0.000	0.000	0.000	0.125	0.000
Northern Russians	83	0.060	0.025	0.018	0.018	0.018	0.024	0.006	0.756	0.024	0.018	0.030	0.120	0.012
Russians of the Southeast of Central Russia	51	0.010	0.000	0.000	0.000	0.000	0.010	0.039	0.622	0.020	0.000	0.029	0.127	0.059
Russians of the Southwest of Central Russia	50	0.020	0.000	0.000	0.000	0.000	0.000	0.010	0.796	0.000	0.000	0.000	0.140	0.010
Russians of the north of the Arkhangelsk region	70	0.029	0.000	0.000	0.000	0.000	0.000	0.000	0.766	0.000	0.000	0.000	0.093	0.000
Siberian Tatars	68	0.007	0.000	0.015	0.015	0.007	0.000	0.000	0.664	0.009	0.000	0.007	0.184	0.000
Tajiks, Pamiris, Yaghnobi people	72	0.014	0.000	0.000	0.000	0.000	0.000	0.000	0.681	0.007	0.000	0.000	0.208	0.000
Tatars	52	0.102	0.160	0.143	0.140	0.130	0.120	0.031	0.717	0.127	0.100	0.135	0.225	0.039
Transcaucasia	77	0.059	0.034	0.040	0.032	0.020	0.019	0.000	0.669	0.027	0.026	0.026	0.125	0.007
Tuvans and Tofalars	55	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.764	0.000	0.000	0.000	0.236	0.000
Ukranians	79	0.019	0.000	0.026	0.000	0.000	0.000	0.000	0.753	0.000	0.000	0.000	0.152	0.000
Uzbeks, Turkmens, Kyrgyz	80	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.700	0.000	0.000	0.000	0.156	0.000
Western Caucasus	87	0.012	0.006	0.000	0.006	0.000	0.000	0.000	0.622	0.000	0.000	0.000	0.149	0.011
Europe	Data from NCBI. gnomAD	0.013	0.000	0.000	0.0005	0.00001	0.0003	0.0004	0.717	0.00	0.00001	0.0006	0.120	0.0001
Africa		0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.571	0.013	0.000	0.076	0.422	0.000
Asia		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.600	0.000	0.010	0.000	0.250	0.030

fathmm web-site [19]. To minimize the number of false positives, a conservative threshold of –3.0 was chosen for analysis. The data obtained were included in Table. 1.

Mathematical and statistical methods

Calculation of population alternative allele frequencies for studied polymorphisms, calculation of χ^2 criterion and *p*-value for assessment of Hardy–Weinberg genotype frequencies, as well as assessment of alternative allele frequency distribution normality in studied populations were performed with RStudio R, version 4.0.2 (RStudio; USA) and Microsoft Excel (Microsoft Corp.; USA). The differences were considered significant at $p < 0.01$.

Multidimensional scaling

Two-dimensional representation of spatial distribution of populations based on the alternative allele frequencies calculated for studied markers *TP53* and *WRAP53* was obtained with STATISTICA10 software package (StatSoft; USA) via multidimensional scaling using the Nei's genetic distances calculated with the DJ genetic software (RCMG; Russia).

RESULTS

Calculation of alternative allele frequencies for studied markers in Russian populations

Based on the genotyping of 28 Russian populations, we calculated alternative allele frequencies for nine exon TP53 polymorphisms (exon 2 — rs201753350; exon 3 — rs1042522, rs1800371; exon 6 — rs144340710; exon 7 — rs112431538, rs149633775; exon 10 — rs587781663, rs17882252, rs150293825), one intronic *TP53* variant (rs17881850), as well as for three polymorphisms of exon 2 of *WRAP53* gene (rs17880282, rs2287499, rs34067256). The calculated population alternative allele frequencies for the listed markers are presented in Table 2.

Alternative allele frequencies and Hardy–Weinberg equilibrium

To test the studied markers *TP53* and *WRAP53* for Hardy–Weinberg equilibrium in the population, χ^2 criterion was calculated based on the existing allele ratio and the calculated in accordance with the Hardy–Weinberg principle marker population frequencies. Table 3 was compiled in order to visualize the relationship between the alternative allele frequencies and the Hardy–Weinberg equilibrium. In five of 28 studied populations (“Central Caucasus”, “Dagestan”, “northern Russians”, “Tatars” and “Transcaucasia”) the combination of high (compared to the listed in Table 2 reference frequencies for the world’s populations of appropriate origin) alternative allele frequencies and their non-equilibrium pattern in the population (orange cells) were observed for most markers. This suggests that the external factors (for example, accidental inbreeding) may affect the pattern of studied alleles in the discussed population. Genotyping errors may also affect the results.

In the population of “southeastern Russians” alternative alleles for many markers were identified; the frequencies of those were higher compared to reference European population, however, the alleles showed no deviation from the Hardy–Weinberg equilibrium.

In some populations (“Komi and Udmurts”, “Siberian Tatars”, “Western Caucasus”), the diversity of identified markers was higher compared to reference populations of appropriate origin, however, their frequencies were low (close to reference values) and satisfied the Hardy–Weinberg principle. Among studied markers, two markers (rs1042522, located in exon 3 of *TP53,* and rs2287499, located in exon 2 of *WRAP53*) were characterized by high frequencies and satisfied the Hardy–Weinberg principle in all populations.

Assessment of normality for distribution of alternative allele frequencies in the populations

Since in theory the neutral alleles are not affected by natural selection, and their population frequencies may follow a normal distribution, we have assessed the normality of the marker frequency distribution in the population using the Shapiro-Wilk test. The test results have made it possible to confirm the null-hypothesis for two markers: rs1042522 ($W = 0.95$, $p = 0.18$) and rs2287499 ($W = 0.97$, $p = 0.46$). For other markers, the normal distribution has not been confirmed.

Analysis of marker population frequencies by multidimensional scaling (MDS)

In our study, multidimensional scaling was the most effective method for the studied populations’ positioning in the low-dimensional space allowing us to evaluate the genetic distances between populations. MDS was performed for 29 populations (Tatar and African populations were excluded due to sharp contrast between their marker frequencies and the data for the main population pool) (see Figure). For the MDS performed the stress value was 0.068, and the alienation coefficient was 0.058.

The populations were pre-labeled as belonging to one of three groups: Asian, European and Caucasian. The populations grouped based on their origin (see Figure) allowed us to determine three appropriate clusters: Asian, European and Caucasian. In the Asian and European clusters having the overlap area of a significant size, the populations are closer to each other based on polymorphism frequencies, and in the Caucasian cluster, the large frequencies’ variability between populations is observed.

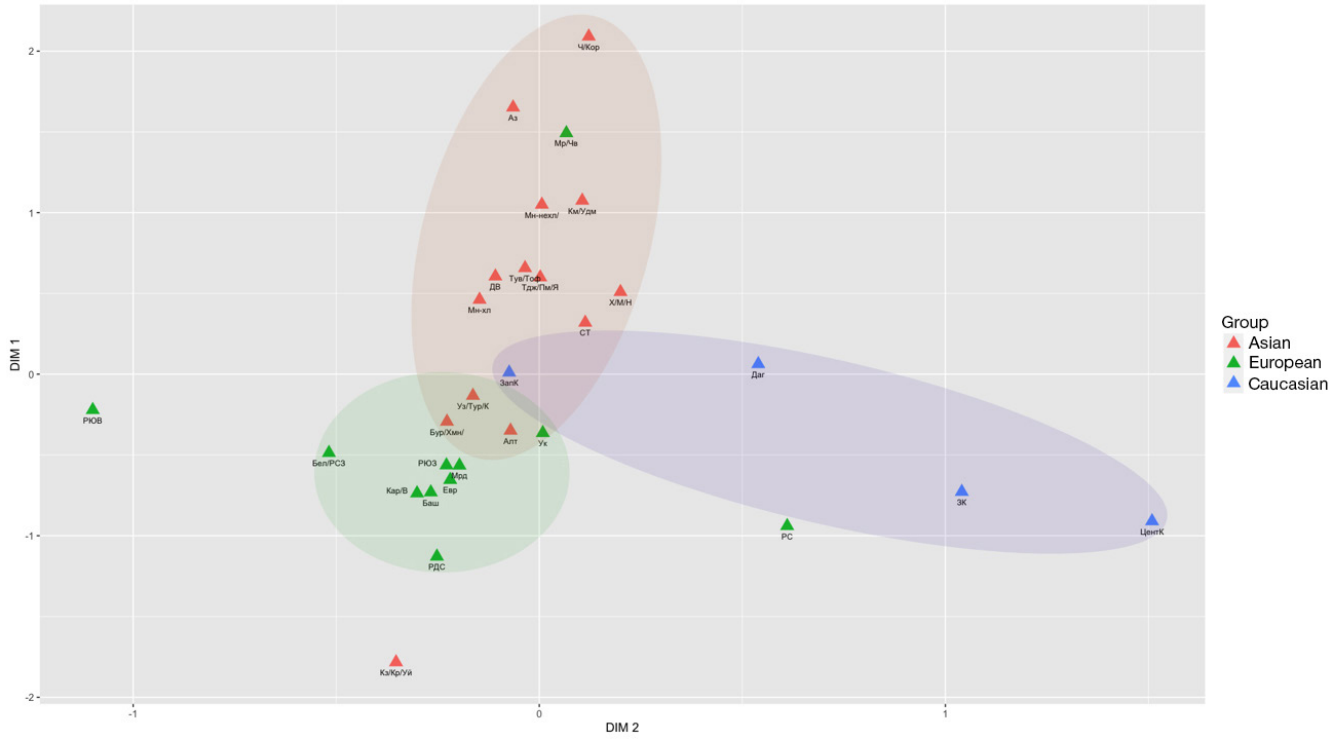
The boundaries of Asian cluster are quite clear. The only exception is the joint population which includes Kazakhs, Karakalpaks, Uighurs and Nogais; the position of this population outside the cluster is due to higher level of some markers showing the equilibrium pattern compared to other Asian populations (Table 3).

European cluster has a more compact shape with high population density around the central reference European population, the marker frequencies for which have been obtained from public sources. The following three European populations appear to be far beyond the cluster: “northern Russians", "southeastern Russians" and the joint population of Mari and Chuvash. The population of “northern Russians” is the only European population showing higher frequencies of many markers deviating from equilibrium, which are absent in other European population (Table 3). Compared to the “northern Russians" population, the population of "southeastern Russians" characterized by high frequencies of some polymorphisms unusual for European populations, the majority of which are in Hardy–Weinberg equilibrium (Table 3), is extended beyond European cluster in the opposite direction on the plot. The "Mari and Chuvash" population is deep inside the Asian cluster, which may be due to anthropological composition

Table 3. Overlapping of marker population frequencies and assessment of Hardy–Weinberg equilibrium

Population	*TP53*										*WRAP53*		
	rs17881850	rs587781663	rs17882252	rs150293825	rs112431538	rs149633775	rs144340710	rs1042522	rs1800371	rs201753350	rs17880282	rs2287499	rs34067256
Altaians	0.026	0	0.007	0	0	0	0	0.766	0	0	0.006	0.156	0
Bashkirs	0	0.01	0	0	0	0	0	0.720	0	0	0	0.116	0
Belarusians and Russians of the Northwest	0.017	0	0	0	0	0	0.017	0.650	0	0	0	0.117	0.033
Buryats, Khamnigans, Yakuts	0	0	0	0	0	0	0	0.772	0	0	0	0.158	0
Central Caucasus	0.071	0.048	0.063	0.047	0.032	0.032	0	0.697	0.032	0.032	0.047	0.125	0.016
Chukchis and Koryaks	0	0	0	0	0	0	0	0.657	0	0	0	0.313	0
Dagestan	0.025	0.013	0.019	0.013	0.013	0.013	0	0.643	0.019	0.025	0.013	0.165	0.025
Far East (peoples of the Amur region)	0	0	0	0	0.006	0	0	0.690	0	0	0.012	0.208	0
Karelians and Vepsians	0.025	0	0.008	0	0	0	0	0.797	0	0	0	0.042	0
Kazakhs, Karakalpaks, Uighurs, Nogays	0.034	0.011	0	0	0	0	0	0.727	0.011	0	0	0.216	0
Khanty, Mansi, Nenets	0	0	0	0	0	0	0	0.717	0	0	0	0.113	0
Komi and Udmurts	0.006	0.006	0.006	0	0.012	0	0	0.645	0.006	0	0.006	0.232	0
Mari and Chuvash	0	0	0	0	0	0	0	0.651	0	0	0	0.264	0.01
Khalkha Mongols	0.010	0	0	0	0	0	0	0.688	0	0	0	0.198	0.02
Mongols (other groups) and Kalmyks	0	0	0	0	0	0	0	0.654	0	0	0	0.231	0
Mordvins	0.013	0	0	0	0	0	0	0.700	0	0	0	0.125	0
Northern Russians	0.060	0.025	0.018	0.018	0.018	0.024	0.01	0.756	0.024	0.018	0.030	0.120	0.012
Russians of the Southeast of Central Russia	0.010	0	0	0	0	0.010	0.039	0.622	0.02	0	0.029	0.127	0.059
Russians of the Southwest of Central Russia	0.020	0	0	0	0	0	0.01	0.796	0	0	0	0.140	0.010
Russians of the north of the Arkhangelsk region	0.029	0	0	0	0	0	0	0.766	0	0	0	0.093	0
Siberian Tatars	0.007	0	0.015	0.015	0.007	0	0	0.664	0.009	0	0.007	0.184	0
Tajiks, Pamiris, Yaghnobi people	0.014	0	0	0	0	0	0	0.681	0.007	0	0	0.208	0
Tatars	0.102	0.160	0.143	0.140	0.130	0.120	0.031	0.717	0.127	0.100	0.135	0.225	0.039
Transcaucasia	0.059	0.034	0.040	0.032	0.020	0.019	0	0.669	0.027	0.026	0.026	0.125	0.007
Tuvans and Tofalars	0.009	0	0	0	0	0	0	0.764	0	0	0	0.236	0
Ukranians	0.019	0	0.026	0	0	0	0	0.753	0	0	0	0.152	0
Uzbeks, Turkmens, Kyrgyz	0.006	0	0	0	0	0	0	0.700	0	0	0	0.156	0
Western Caucasus	0.012	0.006	0	0.006	0	0	0	0.622	0	0	0	0.149	0.011

Note: Hardy–Weinberg equilibrium is assessed based on the χ^2 *p*-value ($p < 0.01$); *orange cells* — null-hypothesis for alternative allele is rejected based on the χ^2 test, *green cells* — null-hypothesis is not rejected.



Алт	Altaians
Баш	Bashkirs
Бел/РСЗ	Belarusians and Russians of the Northwest
Бур/Хмн/	Buryats, Khamnigans, Yakuts
ЦентК	Central Caucasus
Ч/Кор	Chukchis and Koryaks
Даг	Dagestan
ДВ	Far East (peoples of the Amur region)
Кар/В	Karelians and Vepsians
Кз/Кр/Уй	Kazakhs, Karakalpaks, Uighurs, Nogays
Х/М/Н	Khanty, Mansi, Nenets
Км/Удм	Komi and Udmurts
Мр/Чв	Mari and Chuvash
Мн-хл	Khalkha Mongols
Мн-нехл/	Mongols (other groups) and Kalmyks
Мрд	Mordvins
РС	Northern Russians
РЮВ	Russians of the Southeast of Central Russia
РЮЗ	Russians of the Southwest of Central Russia
РДС	Russians of the north of the Arkhangelsk region
СТ	Siberian Tatars
Тдж/Пм/Я	Tajiks, Pamiris, Yaghnobi people
ЗК	Transcaucasia
Тув/Тоф	Tuvans and Tofalars
Ук	Ukranians
Уз/Тур/К	Uzbeks, Turkmens, Kyrgyz
ЗапК	Western Caucasus
Евр	Europe
Аз	Asia

Fig. Multidimensional scaling plot based on Nei's genetic distance matrix for 29 populations (Tatar and African populations were excluded from analysis due to excessively large frequency differences with the rest of populations); stress value is 0.068, and the alienation coefficient is 0.058. The populations are divided into three groups: Asian (*red triangle*), European (*green triangle*) and Caucasian (*blue sign*). The appropriate clusters have been distinguished: *orange* — Asian populations, *green* — European populations, *blue* — Caucasian populations.

of the Chuvash that includes both Caucasian individuals and a significant proportion of Mongoloid individuals together with mixed forms.

All populations of Caucasian cluster, except for the "Western Caucasus" population falling in the overlap area with Asian cluster and being in the close proximity to European cluster, are characterized by high frequencies of all analyzed polymorphisms and their non-equilibrium patterns (Table 3), which brings them close to "northern Russians" based on the discussed parameters (see Figure).

DISCUSSION

The study of *TP53* and *WRAP53* gene polymorphisms in 28 populations covering all major regions of Russia made it possible to assess the frequency and distribution of selected markers in various Russian regions and populations.

Assessment of alternative allele frequencies for the studied markers in Russian populations revealed two major trends:

– two of 13 studied markers (rs1042522, located in exon 3 of *TP53*, and rs2287499, located in exon 2 of *WRAP53*) are characterized by high frequencies in all populations, normal distribution of marker population frequencies, and Hardy–Weinberg proportions of alleles;

– in five populations ("Central Caucasus", "Dagestan", "northern Russians", "Tatars" and "Transcaucasia"), the frequencies of most markers (except the mentioned above rs1042522 and rs2287499, as well as the benign rs144340710 located in exon 6 of TP53) appeared to be high, and the alleles of these markers showed non-equilibrium patterns. In the Tatar population the abundance, frequencies and distribution of the studied polymorphisms' alternative alleles were significantly higher compared to reference global values; they also significantly exceeded the frequencies and abundance of these polymorphisms in the major pool of Russian populations (Table 2). This matter is subject to further research.

Two prevalent equilibrium markers, rs1042522 of *TP53* gene and rs2287499 of *WRAP53* gene, are listed in the ClinVar data base as benign markers. It means that their frequency is too high for markers to be pathogenic mutations; markers are found in hetero- and homozygous state in individuals without severe disease for that gene; markers show no disease association in appropriately sized case-control studies [20]. The fact that the alternative allele of rs1042522 polymorphism being a part of p53 DNA-binding domain [21] shows high frequency in all populations compared to reference genome may indicate the reference genome carrying a random minor allele. This assumption is indirectly supported by the reported increased functional ability to induce apoptosis and prevent cancer development in alternative Arg72 variant of p53 protein compared to reference Pro72 variant [22].

Despite rs1042522 and rs2287499 are listed in scientific databases as benign markers, literature contains a large amount of data on their involvement in carcinogenesis. In particular, it has been reported that heterozygous variant Arg/Pro of p53 Pro72Arg polymorphism (Table 1) is associated with high risk of melanoma compared to heterozygous variant Pro/Pro [23]; another paper reports association of the marker Pro/Pro genotype with increased risk of nonsmall cell lung cancer in patients from Moscow Region [21].

There are many literary sources on survival of cancer patients, homo- and heterozygous for Pro72Arg, however, the reported data is controversial. There is evidence of increased median survival time in cervical cancer patients carrying Arg/Pro genotype when compared with patients with Arg/Arg and Pro/Pro genotypes [24], however, the extensive research carried out by Danish specialists [25] has shown no association of the mentioned above polymorphism with lower mortality after cancer and lower cancer incidence in the general population.

However, rs1042522 is better known as a marker included in the expert panel of Pharmacogenomics Knowledgebase, which is associated with altered body response to some antineoplastic drugs [26]. There is evidence of the p53 Pro allele association with toxicity due to chemotherapy [27], as well as evidence of lower response rate to fluorouracil-based chemotherapy in gastric cancer patients carrying the Pro/Pro genotype compared to patients carrying the Arg/Arg genotype [28].

The clinical significance of rs2287499 marker of gene *WRAP53* is much less well understood, however, there is evidence of moderate linkage disequilibrium between studied markers rs1042522 and rs2287499. The haplotype combination CA/GC is associated with increased risk of breast cancer, and the haplotype combination GA/CC, by contrast, is assumed to be a protective factor against breast cancer [29].

The abundance and frequencies of other polymorphisms in Russian populations vary significantly, however the calculated frequencies of most polymorphisms correspond to reference marker frequencies for Asian and European populations in accordance with the origin of surveyed Russian populations. Exceptions are five listed above populations with high frequencies of the discussed markers. The clinical significance of some studied polymorphisms (for example, the intronic variant rs17881850) remains poorly understood. However, recent evidence suggests that intronic *TP53* gene polymorphisms may also have some clinical significance [30].

CONCLUSION

The study allowed us to obtain data on germline *TP53* (10 markers of five exons and one intron) and *WRAP53* (three markers of exon 2) polymorphism frequencies for 28 Russian populations. In the majority of populations the calculated polymorphism frequencies are close to values obtained for reference global population (Asian or European) of appropriate origin. Six populations ("Central Caucasus", "Dagestan", "northern Russians", "south-eastern Russians", "Tatars" and "Transcaucasia") are characterized by increased marker frequencies compared to reference values; in all listed populations except the population of "southeastern Russians" the marker alleles with high frequencies do not satisfy the Hardy–Weinberg principle. The Tatar population is characterized by high frequencies of polymorphisms' allelic disequilibrium, which demonstrates the need for more detailed investigation of those in this population in order to reveal the cause of such differences.

References

1. Joruiz SM, Bourdon JC. P53 isoforms: Key regulators of the cell fate decision. Cold Spring Harbor Perspectives in Medicine. 2016; 6: 8.
2. Surget S, Khoury MP, Bourdon JC. Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. OncoTargets and Therapy. 2013; 7: 57–68.
3. Rassoolzadeh H. Unwrapping the role of WRAP53β in DNA damage response [Internet]. [Solna]; 2016. Available from: http://ceder.graphics.
4. Mahmoudi S, Henriksson S, Corcoran M, Méndez-Vidal C, Wiman KG, Farnebo M. Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. Molecular Cell. 2009; 33 (4): 462–71.
5. Farnebo M. Wrap53, a novel regulator of p53. Cell Cycle. 2009; 8 (15): 2343–6. DOI: 10.4161/cc.8.15.9223.
6. Henriksson S, Farnebo M. On the road with WRAP53β: guardian of Cajal bodies and genome integrity. Front Genet. 2015; 6: 91. doi: 10.3389/fgene.2015.00091.
7. Bergstrand S, O'Brien EM, Farnebo M. The Cajal body protein WRAP53β prepares the scene for repair of DNA double-strand breaks by regulating local ubiquitination. Front Mol Biosci. 2019; 6: 51. Published 2019 Jul 4. DOI:10.3389/fmolb.2019.00051.
8. Mahmoudi S. WRAP53 unwrapped; roles in nuclear architecture and cancer. 2011.
9. Coucoravas C, Dhanjal S, Henriksson S, Böhm S, Farnebo M. Phosphorylation of the Cajal body protein WRAP53β by ATM promotes its involvement in the DNA damage response. RNA Biol. 2017; 14(6): 804–13. DOI:10.1080/15476286.2016.1243647.
10. Rogoża-Janiszewska E, Malińska K, Górski B, Scott RJ, Cybulski C, Kluźniak W, et al. Prevalence of germline TP53 variants among early-onset breast cancer patients from Polish population. Breast Cancer. 2020.
11. rs2287499 RefSNP Report - dbSNP - NCBI [Internet]. Available from: https://www.ncbi.nlm.nih.gov/snp/rs2287499#frequency_tab.
12. rs2287499 (SNP) — Population genetics — Homo_sapiens — Ensembl genome browser 101 [Internet]. Available from: http://www.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=17:7688350-7689350;v=rs2287499;vdb=variation;vf=87573072
13. rs2287499 | gnomAD v2.1.1 | gnomAD [Internet]. Available from: https://gnomad.broadinstitute.org/variant/rs2287499?dataset=gnomad_r2_1.
14. rs2287499 C>G | Genokarta — geneticheskaja enciklopedija [Internet]. Available from: https://genokarta.ru/snps/rs2287499_CG.
15. Balanovskaya EV, Zhabagin MK, Agdzhoyan AT, Chukhryaeva MI, Markina NV, Balaganskaya OA et al. Populyaczionnye biobanki: princzipy organizaczii i perspektivy primeneniya v genogeografii i personalizirovannoj mediczine. Genetika. 2016; 52 (12): 1371–87.
16. National Center for Biotechnology Information [Internet]. Available from: https://www.ncbi.nlm.nih.gov/.
17. What is ClinVar? [Internet]. Available from: https://www.ncbi.nlm.nih.gov/clinvar/intro/.
18. gnomAD [Internet]. Available from: https://gnomad.broadinstitute.org/.
19. fathmm — Home [Internet]. Available from: http://fathmm.biocompute.org.uk/.
20. Pesaran T, Karam R, Huether R, Li S, Farber-Katz S, Chamberlin A, et al. Beyond DNA: an integrated and functional approach for classifying germline variants in breast cancer genes. Int J Breast Cancer. 2016; 2016: 2469523. DOI: 10.1155/2016/2469523.
21. Zavarykina T, Byrdennyy A, Loginov V, Atkarskaya M, Zhizhina G. A84: Polymorphic markers Arg72Pro and Gln157Lys of TP53 gene in nonsmall cell lung cancer. European Journal of Cancer Supplements [Internet]. 2015; 13 (1): 69. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1359634915001238.
22. Zheltuhin AO, Chumakov PM. Povsednevnye i induciruemye funkcii gena p53. 2010; 50: 447–516.
23. Geng P, Liao Y, Ruan Z, Liang H. Increased risk of cutaneous melanoma associated with p53 Arg72pro polymorphism. PLoS ONE [Internet]. 2015; 10 (3): e0118112. Available from: https://dx.plos.org/10.1371/journal.pone.0118112.
24. Coelho A, Nogueira A, Soares S, Assis J, Pereira D, Bravo I, et al. TP53 Arg72Pro polymorphism is associated with increased overall survival but not response to therapy in Portuguese/Caucasian patients with advanced cervical cancer. Oncology Letters [Internet]. 2018; 15 (5): 8165–71. Available from: http://www.spandidos-publications.com/10.3892/ol.2018.8354/abstract.
25. Kodal JB, Vedel-Krogh S, Kobylecki CJ, Nordestgaard BG, Bojesen SE. TP53 Arg72Pro, mortality after cancer, and all-cause mortality in 105,200 individuals. Scientific Reports. 2017; 7 (1).
26. rs1042522 - Clinical Annotations [Internet]. Available from: https://www.pharmgkb.org/variant/PA166155173/clinicalAnnotation.
27. Henríquez-Hernández LA, Murias-Rosales A, González-Hernández A, de León AC, Díaz-Chico N, Fernández-Pérez L. Distribution of TYMS, MTHFR, p53 and MDR1 gene polymorphisms in patients with breast cancer treated with neoadjuvant chemotherapy. Cancer Epidemiology. 2010; 34 (5): 634–8. DOI: 10.1016/j.canep.2010.06.013.
28. Huang ZH, Hua D, Li LH, Zhu J De. Prognostic role of p53 codon 72 polymorphism in gastric cancer patients treated with fluorouracil-based adjuvant chemotherapy. J Cancer Res Clin Oncol. 2008; 134 (10): 1129–34. DOI: 10.1007/s00432-008-0380-8.
29. Pouladi N, Abdolahi S, Farajzadeh D, Feizi MAH. Haplotype and linkage disequilibrium of TP53-WRAP53 locus in Iranian-Azeri women with breast cancer. Roemer K, editor. PLOS ONE [Internet]. 2019; 14 (8): e0220727. Available from: https://dx.plos.org/10.1371/journal.pone.0220727.
30. Voropaeva EN, Pospeloya TI, Voevoda MI, Mabimovv N. Changes in non-coding sequences of the tp53 gene in diffuse large b-cell lymphoma. Gematologiya i Transfusiologiya. 2018; 63 (3): 239–49.

Литература

1. Joruiz SM, Bourdon JC. P53 isoforms: Key regulators of the cell fate decision. Cold Spring Harbor Perspectives in Medicine. 2016; 6: 8.
2. Surget S, Khoury MP, Bourdon JC. Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. OncoTargets and Therapy. 2013; 7: 57–68.
3. Rassoolzadeh H. Unwrapping the role of WRAP53β in DNA damage response [Internet]. [Solna]; 2016. Available from: http://ceder.graphics.
4. Mahmoudi S, Henriksson S, Corcoran M, Méndez-Vidal C, Wiman KG, Farnebo M. Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. Molecular Cell. 2009; 33 (4): 462–71.
5. Farnebo M. Wrap53, a novel regulator of p53. Cell Cycle. 2009; 8 (15): 2343–6. DOI: 10.4161/cc.8.15.9223.
6. Henriksson S, Farnebo M. On the road with WRAP53β: guardian of Cajal bodies and genome integrity. Front Genet. 2015; 6: 91. doi: 10.3389/fgene.2015.00091.
7. Bergstrand S, O'Brien EM, Farnebo M. The Cajal body protein WRAP53β prepares the scene for repair of DNA double-strand breaks by regulating local ubiquitination. Front Mol Biosci. 2019; 6: 51. Published 2019 Jul 4. DOI:10.3389/fmolb.2019.00051.
8. Mahmoudi S. WRAP53 unwrapped; roles in nuclear architecture and cancer. 2011.
9. Coucoravas C, Dhanjal S, Henriksson S, Böhm S, Farnebo M. Phosphorylation of the Cajal body protein WRAP53β by ATM promotes its involvement in the DNA damage response. RNA Biol. 2017; 14(6): 804–13. DOI:10.1080/15476286.2016.1243647.
10. Rogoża-Janiszewska E, Malińska K, Górski B, Scott RJ, Cybulski C, Kluźniak W, et al. Prevalence of germline TP53 variants among early-onset breast cancer patients from Polish population. Breast Cancer. 2020.

11. rs2287499 RefSNP Report - dbSNP - NCBI [Internet]. Available from: https://www.ncbi.nlm.nih.gov/snp/rs2287499#frequency_tab.
12. rs2287499 (SNP) — Population genetics — Homo_sapiens — Ensembl genome browser 101 [Internet]. Available from: http://www.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=17:7688350-7689350;v=rs2287499;vdb=variation;vf=87573072
13. rs2287499 | gnomAD v2.1.1 | gnomAD [Internet]. Available from: https://gnomad.broadinstitute.org/variant/rs2287499?dataset=gnomad_r2_1
14. rs2287499 C>G | Генокарта — генетическая энциклопедия [Internet]. Available from: https://genokarta.ru/snps/rs2287499_CG.
15. 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.
16. National Center for Biotechnology Information [Internet]. Available from: https://www.ncbi.nlm.nih.gov/.
17. What is ClinVar? [Internet]. Available from: https://www.ncbi.nlm.nih.gov/clinvar/intro/.
18. gnomAD [Internet]. Available from: https://gnomad.broadinstitute.org/.
19. fathmm — Home [Internet]. Available from: http://fathmm.biocompute.org.uk/.
20. Pesaran T, Karam R, Huether R, Li S, Farber-Katz S, Chamberlin A, et al. Beyond DNA: an integrated and functional approach for classifying germline variants in breast cancer genes. Int J Breast Cancer. 2016; 2016: 2469523. DOI: 10.1155/2016/2469523.
21. Zavarykina T, Byrdennyy A, Loginov V, Atkarskaya M, Zhizhina G. A84: Polymorphic markers Arg72Pro and Gln157Lys of TP53 gene in nonsmall cell lung cancer. European Journal of Cancer Supplements [Internet]. 2015; 13 (1): 69. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1359634915001238.
22. Желтухин А. О., Чумаков П. М. Повседневные и индуцируемые функции гена p53. 2010; 50: 447–516.
23. Geng P, Liao Y, Ruan Z, Liang H. Increased risk of cutaneous melanoma associated with p53 Arg72pro polymorphism. PLoS ONE [Internet]. 2015; 10 (3): e0118112. Available from: https://dx.plos.org/10.1371/journal.pone.0118112.
24. Coelho A, Nogueira A, Soares S, Assis J, Pereira D, Bravo I, et al. TP53 Arg72Pro polymorphism is associated with increased overall survival but not response to therapy in Portuguese/Caucasian patients with advanced cervical cancer. Oncology Letters [Internet]. 2018; 15 (5): 8165–71. Available from: http://www.spandidos-publications.com/10.3892/ol.2018.8354/abstract.
25. Kodal JB, Vedel-Krogh S, Kobylecki CJ, Nordestgaard BG, Bojesen SE. TP53 Arg72Pro, mortality after cancer, and all-cause mortality in 105,200 individuals. Scientific Reports. 2017; 7 (1).
26. rs1042522 - Clinical Annotations [Internet]. Available from: https://www.pharmgkb.org/variant/PA166155173/clinicalAnnotation.
27. Henríquez-Hernández LA, Murias-Rosales A, González-Hernández A, de León AC, Díaz-Chico N, Fernández-Pérez L. Distribution of TYMS, MTHFR, p53 and MDR1 gene polymorphisms in patients with breast cancer treated with neoadjuvant chemotherapy. Cancer Epidemiology. 2010; 34 (5): 634–8. DOI: 10.1016/j.canep.2010.06.013.
28. Huang ZH, Hua D, Li LH, Zhu J De. Prognostic role of p53 codon 72 polymorphism in gastric cancer patients treated with fluorouracil-based adjuvant chemotherapy. J Cancer Res Clin Oncol. 2008; 134 (10): 1129–34. DOI: 10.1007/s00432-008-0380-8.
29. Pouladi N, Abdolahi S, Farajzadeh D, Feizi MAH. Haplotype and linkage disequilibrium of TP53-WRAP53 locus in Iranian-Azeri women with breast cancer. Roemer K, editor. PLOS ONE [Internet]. 2019; 14 (8): e0220727. Available from: https://dx.plos.org/10.1371/journal.pone.0220727.
30. Voropaeva EN, Pospeloya TI, Voevoda MI, Mabimovv N. Changes in non-coding sequences of the tp53 gene in diffuse large b-cell lymphoma. Gematologiya i Transfusiologiya. 2018; 63 (3): 239–49.