

DATA NOTE

Open Access



Genotype data for 60 SNP genetic markers associated with eye, hair, skin color, ABO blood group, sex, core Y-chromosome haplogroups in Kazakh population

Alizhan Bukayev¹, Baglan Aidarov¹, Denis Fesenko², Viktoriya Saidamarova³, Ivan Ivanovsky⁴, Elina Maltseva^{5,6}, Dinara Naizabayeva⁵, Ayagoz Bukayeva¹, Bekzhan Faizov¹, Vladimir Pylev⁷, Akynkali Darmenov³, Yuriy Skiba⁵, Elena Balanovska⁷ and Maxat Zhabagin^{1*}

Abstract

Objectives The collection of genotype data was conducted as an essential part of a pivotal research project with the goal of examining the genetic variability of skin, hair, and iris color among the Kazakh population. The data has practical application in the field of forensic DNA phenotyping (FDA). Due to the limited size of forensic databases from Central Asia (Kazakhstan), it is practically impossible to obtain an individual identification result based on forensic profiling of short tandem repeats (STRs). However, the pervasive use of the FDA necessitates validation of the currently employed set of genetic markers in a variety of global populations. No such data existed for the Kazakhs. The Phenotype Expert kit (DNA Research Center, LLC, Russia) was used for the first time in this study to collect data.

Data description The present study provides genotype data for a total of 60 SNP genetic markers, which were analyzed in a sample of 515 ethnic Kazakhs. The dataset comprises a total of 41 single nucleotide polymorphisms (SNPs) obtained from the HlrisPlex-S panel. Additionally, there are 4 SNPs specifically related to the ABO gene, 1 marker associated with the AMELX/Y genes, and 14 SNPs corresponding to the primary haplogroups of the Y chromosome. The aforementioned data could prove valuable to researchers with an interest in investigating genetic variability and making predictions about phenotype based on eye color, hair color, skin color, ABO blood group, gender, and biogeographic origin within the male lineage.

Keywords Forensic DNA Phenotyping (FDA), SNP Genotyping, Microarray, Eye color, Hair color, Skin color, ABO blood group, Y-chromosome haplogroup, Sex, Kazakh population

*Correspondence:

Maxat Zhabagin
mzhabagin@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Objective

Genotype data was collected as part of a pivotal study to study the genetic variability of skin, hair, and iris color in the Kazakh population. The aim of the pivotal study is to evaluate the accuracy of predicting the color of the skin, hair, and iris using the genetic markers of the HIRISplex-S set [1]. Genotype data has been published for the first time. This data will be integrated with phenotypic data for analysis in a forthcoming article, according to the purpose of the pivotal study.

The Kazakh population is classified as belonging to a mixed anthropological type known as the South Siberian race. This race can be identified as a transitional group that emerged from the intermingling of Mongoloid and Caucasoid populations in the southern regions of Siberia, Kazakhstan, and Central Asia [2]. The geographic range inhabited by the Kazakh population spans from the western banks of the Volga River to the eastern slopes of the Altai Mountains. It extends from the northern West Siberian lowland to the southern Kyzylkum desert and the Tien Shan mountain system. This vast territory ranks ninth globally in terms of its expanse. Hence, it is reasonable to anticipate a considerable range of phenotypic variations in facial characteristics within the Kazakh population. Nevertheless, it is worth noting that the phenotypic characteristics associated with light skin Kazakh individuals may possess unique genetic underpinnings in terms of pigmentation. This is due to the fact that communities residing in the border regions between Asia and Europe exhibit genetic variations that distinguish them from European populations [3].

The study has applications in several fields: medicine (pigmentation pathology) [4], physical anthropology and archeology (reconstruction of appearance from ancient DNA) [5, 6], as well as forensic DNA phenotyping (FDA) [7, 8]. Due to the limited size of forensic databases from Central Asia (Kazakhstan), it is practically impossible to obtain an individual identification result based on forensic profiling of short tandem repeats (STRs). Consequently, the latter field is particularly pertinent in situations where it is impossible to obtain an individual identification result. Therefore, the FDA is developing swiftly today, including the prediction of a person's externally visible characteristics in relation to their appearance, biogeographic origin, and age based on their DNA [9]. However, their widespread use in forensic practice requires validation of the proposed sets of genetic markers in different populations around the world. There was no such data for the Kazakhs. In this study, data was obtained for the first time using the Phenotype Expert kit (DNA Research Center, LLC, Russia) [10].

Data description

Genotype data was obtained for 515 individuals from the Kazakh population. Individuals were selected from volunteers who took part in this study. Each individual familiarized himself with the aims and methods of the study, signed an informed consent for his/her participation, and provided 10 ml of venous blood for the study. The Local Ethical Commission at the National Center for Biotechnology (Kazakhstan) has previously reviewed and approved procedures relating to blood sampling and other research methods (No. 3 of August 7, 2020). The criteria for random selection included individuals who had resided in Kazakhstan for at least three generations, with the exclusion of cousins within three generations. The sample under study comprises 162 female participants and 353 male participants, with mean ages of 22 years for females and 21 years for males.

The isolation of DNA from venous blood was conducted using the Wizard (R) Genomic DNA Purification Kit (Promega, USA) in accordance with the protocol provided by the manufacturer. The concentration of DNA was measured using a NanoDrop One instrument (Thermo Fisher Scientific, USA) and a Quantus Fluorometer (Promega, USA) with the QuantiFluor(R) ONE dsDNA System kit (Promega, USA) following the instructions provided by the manufacturer. Genotyping of markers associated with eye color, hair color, skin color, AB0 blood group, basic Y-chromosome DNA haplogroup, and gender was carried out using a Phenotype Expert kit (DNA Research Center, LLC, Russia). A test included multiplex PCR, hybridization of the PCR product on a biochip, and genotype determination. Multiplex PCR was performed to amplify and fluorescently label 53 target fragments contained 60 genetic markers associated with eye color, hair color, skin color, AB0 blood group, basic Y-chromosome DNA haplogroup and sex. The detailed description of Phenotype Expert kit has been provided previously [10]. PCR was performed in 25 μ l of the following composition: PCR buffer with HotTaqMulti polymerase, 4 units (Asfogen, Russia), 5 mM MgSO₄, 0.2 mM of each of the dNTPs (Sibenzym, Russia), primer mixture, 200 pmol of Cy5-TCATTG GATCTCATT universal primer, 5 ng of genomic DNA. Amplification was carried out in 0.2 ml PCR-tubes on a CFX96 thermocycler (Bio-Rad, USA) with the following conditions: 95 °C for 2 min and 50 cycles in the first stage (95 °C for 20 s, 65 °C for 30 s, 66 °C for 30 s, 69 °C for 40 s), then 40 cycles in the second stage (95 °C for 20 s, 56 °C for 30 s, 72 °C for 30 s). The 30- μ l chamber of the biochip was filled with a mixture of the following composition: 7.5 μ l formamide, 7.5 μ l 20 \times SSPE, 15 μ l PCR product. After incubation (12 h, 37 °C) and washing (10 min in 1 \times SSPE at room temperature), the biochips

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Raw genotype data for Kazakh population	VCF(vcf.gz)	dbSNP (https://identifiers.org/dbsnp:rs1042602) [11]
			dbSNP (https://identifiers.org/dbsnp:rs1393350) [11]
			dbSNP (https://identifiers.org/dbsnp:rs1126809) [11]
			dbSNP (https://identifiers.org/dbsnp:rs12821256) [11]
			dbSNP (https://identifiers.org/dbsnp:rs12896399) [11]
			dbSNP (https://identifiers.org/dbsnp:rs2402130) [11]
			dbSNP (https://identifiers.org/dbsnp:rs17128291) [11]
			dbSNP (https://identifiers.org/dbsnp:rs1545397) [11]
			dbSNP (https://identifiers.org/dbsnp:rs1800414) [11]
			dbSNP (https://identifiers.org/dbsnp:rs1800407) [11]
			dbSNP (https://identifiers.org/dbsnp:rs12441727) [11]
			dbSNP (https://identifiers.org/dbsnp:rs1470608) [11]
			dbSNP (https://identifiers.org/dbsnp:rs1129038) [11]
			dbSNP (https://identifiers.org/dbsnp:rs12913832) [11]
			dbSNP (https://identifiers.org/dbsnp:rs2238289) [11]
			dbSNP (https://identifiers.org/dbsnp:rs6497292) [11]
			dbSNP (https://identifiers.org/dbsnp:rs1667394) [11]
			dbSNP (https://identifiers.org/dbsnp:rs1426654) [11]
			dbSNP (https://identifiers.org/dbsnp:rs3114908) [11]
			dbSNP (https://identifiers.org/dbsnp:rs3212355) [11]
			dbSNP (https://identifiers.org/dbsnp:rs312262906) [11]
			dbSNP (https://identifiers.org/dbsnp:rs1805005) [11]
			dbSNP (https://identifiers.org/dbsnp:rs1805006) [11]
			dbSNP (https://identifiers.org/dbsnp:rs2228479) [11]
			dbSNP (https://identifiers.org/dbsnp:rs11547464) [11]
			dbSNP (https://identifiers.org/dbsnp:rs1805007) [11]
			dbSNP (https://identifiers.org/dbsnp:rs201326893) [11]
			dbSNP (https://identifiers.org/dbsnp:rs1110400) [11]
			dbSNP (https://identifiers.org/dbsnp:rs1805008) [11]
			dbSNP (https://identifiers.org/dbsnp:rs885479) [11]
dbSNP (https://identifiers.org/dbsnp:rs1805009) [11]			
dbSNP (https://identifiers.org/dbsnp:rs8051733) [11]			
dbSNP (https://identifiers.org/dbsnp:rs6059655) [11]			
dbSNP (https://identifiers.org/dbsnp:rs6119471) [11]			
dbSNP (https://identifiers.org/dbsnp:rs2378249) [11]			

Table 1 (continued)

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs16891982) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs28777) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs12203592) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs12203592) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs683) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs10756819) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs10756819) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs8176741) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs8176720) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs8176719) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs1717135119) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs35284970) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs34820858) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs9786534) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs2740980) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs2032607) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs563604826) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs202084622) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs9341278) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs2032678) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs2319818) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs17307398) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs576940616) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs2032623) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs9341308) [11]
			<i>dbSNP</i> (https://identifiers.org/dbsnp:rs13447352) [11]

were washed with distilled water, dried with compressed air, placed in a portable analyzer “Picodetect” (OOO BIOCHIP-IMB, Moscow) and fluorescence was registered with an exposure of 0.5–2 s according to the manufacturer’s manual. Image analysis was performed using ImaGel Studio software (EIMB RAS, Moscow). The result of automatic processing of each biochip was a text file with the genotype of the sample. Results obtained for the ABO blood group, basic Y-chromosome haplogroup, and gender were interpreted using the Phenotype Expert

program (DNA Research Center, LLC). The raw genotype data for the 60 genetic markers in the 515 Kazakhs are shown in Additional file 1: Table S1 and Table 1 [11]. We have submitted the data to the National Center for Biotechnology Information Reference Assembly dbSNP repository. The data is publicly accessible in the SNP Submission Batch, which can be viewed via the following link: (https://www.ncbi.nlm.nih.gov/SNP/snp_viewBatch.cgi?sbid=1063555).

Limitations

- Compared to comprehensive genome-wide data, the genotypes given for 60 genetic markers may not show all of the genetic variation in eye color, hair color, skin color, ABO blood type, gender, and the main Y-chromosome haplogroups in the Kazakh population.
- The limited size of the study sample (N = 515 individuals) drawn from the Kazakh population (16 million) may provide limitations in accurately evaluating the association between the examined SNPs and variations in eye, hair, and skin color.

Abbreviations

DNA	Deoxyribonucleic acid
FDA	Forensic DNA Phenotyping
STR	Short Tandem Repeat
SNP	Single nucleotide polymorphisms
dNTPs	Deoxynucleotide triphosphates
PCR	Polymerase chain reaction
SSPE	Sodium Chloride–Sodium Phosphate–EDTA

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13104-024-06712-z>.

Additional file 1. Genotype data for 60 genetic markers associated with eye, hair, skin color, ABO blood group, sex, core Y-chromosome haplogroups in Kazakh population

Acknowledgements

We gratefully acknowledge all sample donors who participated in this study. We thank Mashzhan Akzhigit, Malika Dosmagambetova and Assiya Taizhanova for the research assistance. This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (IRN No. AP09058276).

Author contributions

MZ and EB originally conceptualized the study; The current study was designed by VS, AD, YS and MZ. Collected samples and Data Curation: BA, VP, AD; Performed the experiments: AB1, EM, DN, AB2, BF; Validation: DF, II, YS and VP. Formal analysis: AB1, AB2, BF. Contributed reagents/materials/analysis tools: DF, II, YS, AD and EB; The manuscript was originally drafted by AB1 and BA, corrected by DF and VS, and the final draft was prepared by MZ; Supervision: EB. Funding acquisition: MZ. The manuscript was revised, amended, and approved by all authors.

Funding

This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (No. AP09058276).

Availability of data and materials

The data described in this Data note can be freely and openly accessed on the National Center for Biotechnology Information Reference Assembly dbSNP repository under SNP Submission Batch [https://www.ncbi.nlm.nih.gov/SNP/snp_viewBatch.cgi?sbid=1063555]. Please see Table 1 and references [11] for details and links to the data.

Declarations

Ethics approval and consent to participate

Written, informed consent from all study participants and ethical clearance to conduct this study was obtained from the ehe Local Ethical Commission at the National Center for Biotechnology of Kazakhstan (№3 of 7 August 2020).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹National Center for Biotechnology, Astana 010000, Republic of Kazakhstan. ²Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow 119991, Russia. ³Karaganda Academy of the Ministry of Internal Affairs of the Republic of Kazakhstan Named After Barimbek Beisenov, Karaganda 100000, Republic of Kazakhstan. ⁴DNA Research Center, LLC, Khimki 141402, Russia. ⁵Almaty Branch of the National Center for Biotechnology, Almaty 050054, Kazakhstan. ⁶Tethys Scientific Society, Almaty 050063, Kazakhstan. ⁷Bochkov Research Centre of Medical Genetics, Moscow 115522, Russia.

Received: 25 August 2023 Accepted: 1 February 2024

Published online: 18 February 2024

References

1. Chaitanya L, Breslin K, Zuñiga S, Wirken L, Pośpiech E, Kukla-Bartoszek M, Sijen T, Knijff P, Liu F, Branicki W, Kayser M, Walsh S. The HirisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Sci Int Genet.* 2018;35:123–35. <https://doi.org/10.1016/j.fsigen.2018.04.004>.
2. Alekseev VP. Geography of human races // Izbr. in 5 volumes. T. 2. Anthropogeography. M., 2007. (in russian)
3. Pagani L, Lawson DJ, Jagoda E, Mörseburg A, Eriksson A, Mitt M, Clemente F, Hudjashov G, DeGiorgio M, Saag L, Wall JD, Cardona A, Mägi R, Wilson Sayres MA, Kaewert S, Inchley C, Scheib CL, Järve M, Karmin M, Jacobs GS, Metspalu M. Genomic analyses inform on migration events during the peopling of Eurasia. *Nature.* 2016;538(7624):238–42. <https://doi.org/10.1038/nature19792>.
4. Michaud, V., Lasseaux, E., Green, D. J., Gerrard, D. T., Plaisant, C., UK Biobank Eye and Vision Consortium, Fitzgerald, T., Birney, E., Arveiler, B., Black, G. C., & Sergouniotis, P. I. The contribution of common regulatory and protein-coding TYR variants to the genetic architecture of albinism. *Nat Commun.* 2022;13(1):3939. <https://doi.org/10.1038/s41467-022-31392-3>.
5. Zupanič Pajnič I, Leskovaar T, Črešnar M. Eye and hair color prediction of an early medieval adult and subadult skeleton using massive parallel sequencing technology. *Int J Legal Med.* 2023;137(5):1629–38. <https://doi.org/10.1007/s00414-023-03032-y>.
6. Fesenko DO, Aramova OY, Vdovchenkov EV, Ivanovsky ID, Fesenko OE, Polyakov SA, Faleeva TG, Filippova MA, Florinskaya VS, Kornienko IV. *Mol Biol.* 2023;57(4):597–608.
7. Kayser M, de Knijff P. Improving human forensics through advances in genetics, genomics and molecular biology. *Nat Rev Genet.* 2011;12(3):179–92. <https://doi.org/10.1038/nrg2952>.
8. Balanovska E, Lukianova E, Kagazezheva J, Maurer A, Leybova N, Agdzhoian A, Gorin I, Petrushenko V, Zhabagin M, Pylev V, Kostryukova E, Balanovsky O. Optimizing the genetic prediction of the eye and hair color for North Eurasian populations. *BMC Genomics.* 2020;21(Suppl 7):527. <https://doi.org/10.1186/s12864-020-06923-1>.
9. Kayser M, Branicki W, Parson W, Phillips C. Recent advances in Forensic DNA Phenotyping of appearance, ancestry and age. *Forensic Sci Int Genet.* 2023;65: 102870. <https://doi.org/10.1016/j.fsigen.2023.102870>.
10. Fesenko DO, Ivanovsky ID, Ivanov PL, Zemskova EY, Agapitova AS, Polyakov SA, Fesenko OE, Filippova MA, Zasedatelev AS. *Molekuliarnaia biologiiia. Molrus.* 2022;56(5):860–80. <https://doi.org/10.31857/S0026898422050056>.

11. The National Center for Biotechnology Information Reference Assembly dbSNP repository [https://www.ncbi.nlm.nih.gov/SNP/snp_viewTable.cgi?handle=LHG]

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.