Novel genes and variants associated with longevity in Bulgarian centenarians revealed by whole exome sequencing DNA pools: a pilot study

Dimitar Serbezov*, Lubomir Balabanski*, Sena Karachanak-Yankova, Radoslava Vazharova, Desislava Nesheva, Zora Hammoudeh, Rada Staneva, Marta Mihaylova, Vera Damyanova, Olga Antonova, Dragomira Nikolova, Savina Hadjidekova, Draga Toncheva

Department of Medical Genetics, Medical University of Sofia, Sofia 1431, Bulgaria.

Correspondence to: Dr. Draga Toncheva, Department of Medical Genetics, Medical University of Sofia, Zdrave 2, Sofia 1431, Bulgaria. E-mail: dragatoncheva@gmail.com

How to cite this article: Serbezov D, Balabanski L, Karachanak-Yankova S, Vazharova R, Nesheva D, Hammoudeh Z, Staneva R, Mihaylova M, Damyanova V, Antonova O, Nikolova D, Hadjidekova S, Toncheva D. Novel genes and variants associated with longevity in Bulgarian centenarians revealed by whole exome sequencing DNA pools: a pilot study. J Transl Genet Genom 2020;4:446-54. http://dx.doi.org/10.20517/jtgg.2020.41

Abstract

Aim: To determine specific genetic loci that might be associated with longevity in Bulgarian population by analyzing exome pool-seq data from centenarians and a control group.

Methods: We performed whole-exome sequencing of two DNA pools, set up of 32 Bulgarian centenarians and 61 young healthy controls, respectively, and 59935 quality filtered variants were concurrently detected in both pools. Fisher’s exact test was employed to establish the significance of allele frequency difference between the pools.

Results: Forty seven variants showed significantly higher allele frequency in the centenarian compared to the control pool, and these can be considered to be positively associated with longevity in Bulgarian population. Based on their assigned functional role, three genes containing three of these variants were further investigated. These genes, RNF43, WNK1 and NADSYN1, are involved in evolutionary conserved processes with well ascertained association with longevity, i.e., Wnt signaling pathway, insulin/IGF-1 signal pathway and redox balancing processes, respectively. For the remaining genes exhibiting variants with significantly higher allele frequency in the
Bulgarian centenarian pool there is not enough evidence about their functional role in determining longevity and further research is needed.

**Conclusion**: The results confirm the importance of studying centenarians in different populations to discover those combinations of variants that associate with longer health span.

**Keywords**: Centenarian, exome, Bulgarian

**INTRODUCTION**

The European Union is currently undergoing significant demographic change. The average life expectancy for both men and women continues to increase while the birth rate is declining\(^{[1]}\). As a result of these trends European countries have presently attained the lowest birth rates and among the highest life expectancy rates in the world. The European Commission predicts that this trend will continue over the next few decades, resulting in a significant increase in the proportion of retired people in the population and a concordant decrease in the proportion of working-age individuals. After 50 years, the ratio between workers and retirees is expected to fall from 4:1 to 2:1. Such population aging will unavoidably have consequences on the economic and health sectors. Bulgaria is no exception to these demographic trends, and it is ranked number seven on the list of countries with the highest proportion of people over the age of 65, after Japan, Italy, Greece, Germany, Portugal and Finland. According to the National Statistical Institute at the end of 2015, 20.4% of the Bulgarian population was over this age, and this percentage is increasing each year\(^{[2]}\).

Aging is a genetic process that leads to a decline in the ability of the human body to maintain homeostatic balance. It is one of the most significant risk factors for a number of diseases as it leads to progressive degeneration of tissues and organs. The biological mechanisms behind the complex aging process are not yet fully understood, in humans or in model organisms. Studies of long-lived twins, including centenarian siblings, have shown that longevity has a strong genetic component\(^{[3]}\). However, still only a handful of genetic factors have repeatedly been shown to be associated with extreme longevity\(^{[4,5]}\).

Genetic association studies in centenarians from different populations have revealed that a variety of metabolic, cellular and tissue maintenance mechanisms influence aging\(^{[6,7]}\). Longevity has been shown to be affected by variations in the sequence or expression of genes related to the preservation of telomere length, DNA damage repair, tolerance to stress and heat-shock response, as well as the degree of accumulation or restriction of free radicals. In addition, clinical trials on centenarians have found that variation in genes involved in lipoprotein metabolism (e.g., APOE and APOB), cardiovascular homeostasis, immunity, and inflammatory processes may also contribute to prolonging life and preventing diseases.

Centenarians are individuals who, having attained the full extent of human longevity, present a unique opportunity to gain medical and genetic insight. Their genome is presumably devoid of pathological variants with significant penetrance and/or it carries protective alleles against various environmental risk factors. Centenarian genome can thus be considered the “gold standard” for establishing genetic factors predisposing to longevity. Whole-genome sequencing could facilitate uncovering some of these factors, and it might even unravel the physiological mechanisms impacting longevity. Decoding the molecular mechanisms that govern ageing could facilitate the development of strategies and therapies for prolonging healthy life expectancy, and a key aspect of that is also the preservation of cognitive abilities and physical activity. Extending the working age has major social implications as older people acquire the opportunity to be an active and contributing part of the society.
Ageing is a complex process regulated by diverse physiological mechanisms and longevity is determined by the combined effect of multiple genetic and environmental factors. A multifactorial approach is needed to analyze the complex network of interactions between different genes, as well as the contribution of each gene to longevity. Populations differ in genomic characteristics, and it is conceivable that centenarians in America could have different adaptive mechanisms than centenarians in Europe. Studying the genome of long-lived individuals at population level will contribute to more in-depth understanding of genetic factors related to health and longevity.

Population frequency of centenarians is very low on a worldwide basis (0.006%) and there are marked differences between countries, e.g., 0.002% in India and as high as 0.048% in Japan. In Bulgaria, their frequency is 0.0036%[8].

The cost-effectiveness of sequencing pools of individuals (Pool-seq) provides the basis for the popularity and wide-spread use of this method for many research questions, e.g., unravelling the genetic basis of complex traits. Pool-seq methods have been repeatedly shown to provide reliable estimates of allele frequencies[9]. Pool-seq also has the great advantage of allowing flexible re-analysis of new genes from the same data without the need for repeated sequencing.

The aim of the present study is to determine specific genetic loci that might be associated with longevity by analyzing exome pool-seq data of Bulgarian centenarians and a control group.

METHODS

The study was approved by the Ethics committee of the Medical University of Sofia and was found to be in accordance with the requirements of national and international legislation for conducting research involving human subjects. To ensure full compliance with the principles of information regulations, personal data protection and the right to privacy, participants received prior information about the aim, objectives and the methods used in the project, e.g., sampling, data analyses, etc., so that they can get acquainted in detail and make informed decision about taking part in the project. The subjects were also required to sign consent before biological samples were taken.

Tissue samples (saliva or blood sample) were collected from a group of 32 unrelated Bulgarian centenarians (100-106 years old) from different geographical regions, selected to be capable of walking independently after the age of 90. The control group was set up of 61 young healthy individuals (25-30 years old) ethnically matched but unrelated, neither to each other or to the subjects in the centenarian group. The subjects in this study were selected in such a way as to minimize the effects of the environment, and potential population stratification or admixture. The control subjects have survived the prevailing children's diseases, yet the probability that they will reach extreme longevity is low as they are still susceptible to diseases due to their lifestyle, and these affecting middle and later age individuals. The clinical team, with the assistance of gerontologists, designed the questionnaire in order to gather information about subjects' lifestyle, medical history, neurological status, movement independence, cardiovascular disease, cancer, diabetes, etc. The questionnaire included questions on their nutrition, tobacco and alcohol consumption, physical activity, alongside other potentially significant factors, e.g., social contacts, positive mood, stress periods experienced, financial problems, presence of long-lived family members.

DNA was extracted using QIAamp DNA Blood Mini Kit (Qiagen) and equimolar amounts of DNA were used to prepare the two pools. These were whole exome sequenced using BGI v4 chemistry on a BGISEQ-500 platform (by BGI Genomics) at a mean 250x coverage. Such high coverage is required for pool-seq sequencing to ensure that alleles with low frequency are also detected[10]. Variant calling was performed using GATK and the obtained VCF files were annotated using the web-based platform
Following the ‘best practice’ recommendations for pool-seq data\[9\], we performed robust filtering on variant calling: genotype quality $\geq 99$, mapping quality $\geq 60$, number of reads per MAF $> 2$, total depth of coverage above 30 and below 500. The number of variants annotated simultaneously in both pools after applying these filters was 59,935 ($52,870$ SNPs and $7,065$ indels). The number of allele reads for each variant was used to construct contingency tables and the Fisher’s exact test was then deployed to evaluate the significance of the allele frequency differences. The allele frequency estimates obtained were compared with values taken from the publicly available resource for exome sequencing data, The Genome Aggregation Database (gnomAD)\[12\]. False discovery rate (FDR) adjustment of Benjamini and Hochberg\[13\] was used to reduce the number of false positives. All statistical analyses were performed using R scripts\[14\].

The following criteria were used to identify variants that are likely to be positively associated with longevity in Bulgarian centenarians: (1) significantly higher estimated frequency in centenarians compared to the control group; (2) prioritization of genes accommodating the variant according to their molecular function, e.g., being a part of signaling network, evidence for association with longevity in humans or model organisms; (3) evidence for interaction with variants in other genes known to be associated with longevity; and (4) the impact of specific nucleotid change according to various software algorithms.

RESULTS

Centenarian lifestyle, living environment and medical history

Only a small proportion of the sampled Bulgarian centenarians follow vegetarian diet (2%), 84% consume salty foods, 86% sugar containing and 40% animal fat containing foods. They report that they consume fish, albeit only occasionally, in contrast to the centenarians in the Mediterranean region. Overall, the diet of the Bulgarian centenarians is similar to the typical diet of the country’s population. They are also generally of normal body habitus, corroborating the role of excess body fat as a risk factor for longevity. As much as 76% report, whereas the remaining 24% are uncertain about, the presence of long-living family members. These data support the significance of genetic factors in determining human longevity. All centenarians are with well preserved memory function, and 93% state positive life attitude. Seven percent report that they are occasional smokers, in contrast to the country’s general population where 30% are regular smokers\[15\], and 63% state that they refrain from consuming alcohol, or do so only occasionally. All interviewed centenarians claim that they maintain moderate (62%), and even high (38%), physical activity that includes agricultural or domestic activities, sports, long walks, etc. This also contrasts with the general population where 40% report moderate and only 9% with high physical activity, respectively\[15\].

Whole exome pool-seq data

As an initial step in our analysis, the correlation coefficient between the estimated allele frequencies of the detected variants was calculated between the two pools, Pearson’s $r = 0.89$, as well as between the Bulgarian control pool and estimates for non-Finnish European population, Pearson’s $r = 0.93$ [Figure 1].

Fisher’s exact test identified a number of variants that differ significantly between the Bulgarian centenarian and control pools, and these are visualized on a Manhattan [Figure 2].

Of the 91 variants estimated to differ most significantly between the two pools, those above the line (FDR adj. $P$-value $= 5.0 \times 10^{-5}$) [Figure 2], 47 had significantly higher frequency in the centenarian pool (black dots above the identity line on Figure 3, Supplementary Table 1) and 44 in the control pool (blank dots below the identity line on Figure 3).

It can be speculated that variants found to have significantly higher frequency in the centenarian pool are positively associated with longevity, whereas variants found to have significantly lower frequency in the centenarian pool are negatively associated with longevity in the Bulgarian population.
We used publicly available data bases and performed literature survey in order to establish the functional role of the genes with variants positively associated with longevity. As a result, three variants were nominated to be putatively associated with longevity: rs2526374 in the RNF43 gene, rs956868 in the WNK1 gene and rs2276362 in the NADSYN1 gene [Table 1].
Figure 3. Plot showing allele frequencies of variants significantly differing (FDR adj. P-value < 5.0 × 10^{-8}) between Bulgarian centenarians and the control group (n = 91). The filled circles above the identity line are variants estimated to be in higher frequency in centenarians, the blank circles below the identity line are variants estimated to be in higher frequency in the control group. FDR: false discovery rate

| Gene  | Chr | Ref > Alt | dbSNP       | Function | Exonic function | Alt allele freq in Cont. pool | Alt allele freq in Centen. pool | FDR adj. P-value |
|-------|-----|-----------|-------------|----------|-----------------|-------------------------------|-------------------------------|-----------------|
| RNF43 | 17  | G>T       | rs2526374   | Exonic   | non-synonymous  | 0.32                          | 0.55                          | 4.83E-09        |
| WNK1  | 12  | A>C       | rs956868    | Exonic   | non-synonymous  | 0.65                          | 0.87                          | 2.15E-10        |
| NADSYN1 | 11 | A>G       | rs2276362   | Intronic |                 | 0.47                          | 0.72                          | 2.20E-10        |

Table 1. Variants with significantly higher frequency in centenarians, selected as likely being positively associated with longevity in Bulgarian centenarians

DISCUSSION

In this study, whole exome sequencing was performed on two DNA pools, one set up with Bulgarian centenarians and the other with young and healthy individuals serving as a control group. As an initial step in our analyses we compared the allele frequency estimates from the Bulgarian control group with estimates from the non-Finnish European population[^12]. Besides testifying for the reliability of the pool-seq method used, the high correlation coefficient of allele frequencies between the two groups demonstrates that there is a considerable proportion of variants that markedly differ in frequency between the two populations. This inter-population heterogeneity calls for the need to establish population-specific allele frequency databases that centenarian exomes can be compared to.

Of approximately sixty thousand variants detected in both pools, 91 variants differed significantly in allele frequency between the two pools. Among them 47 variants in 43 genes were with higher frequency in
the centenarian pool and could therefore putatively be associated with extreme human longevity, acting along and in combination with environmental factors in the Bulgarian population. These variants are with estimated high population frequencies [Supplementary Table 1], as is often the case with variants associated with complex traits since such traits are formed by many genetic factors with small effect. Of these 43 genes, only one (GSTZ1) is the listed in LongevityMap database, a comprehensive online database of longevity associated genes and variants. In order to establish common molecular mechanisms for these 43 genes, we performed pathway analyses using ToppGene and Reactome platforms. We could however not establish any common significant pathways associated with this set of genes. Based on their known molecular function established from literature survey, we nominated three variants in genes linked to familiar genetic pathways for longevity. Below we discuss each of these genes and variants, and consider the mechanisms they might be involved in that result in extending longevity in Bulgarian centenarians.

**RNF43 (rs2526374)**
The protein encoded by this gene is thought to negatively regulate Wnt signaling, an important, evolutionary conserved molecular pathway, involved in cell proliferation, tissue homeostasis and maintenance of stem cells in adults. As organisms age, proliferating cells stop dividing and stem cells are lost, but Wnt signaling counteracts this by maintaining stem cells in their niches with positive effects on neurogenesis and bone regeneration. Mutations in the Wnt signaling pathway have been shown to lead to a variety of developmental defects in animals. Mutations in the RNF43 gene have been reported in multiple tumor cells, including colorectal and endometrial cancers. Identifying evolutionarily conserved genes, pathways and mechanisms involved in regulating lifespan and life history is a central goal of aging research and has direct implications to human health, not least because most mechanistic aging research is carried out in model organisms.

**WNK1 (rs956868)**
The protein encoded by this gene is a serine/threonine kinase, playing a major role in cell signaling, proliferation and cell survival. This gene is a downstream effector of the insulin/IGF-1 signal pathway that has been shown to regulate longevity in various model organisms. WNK1 has also been shown to regulate the activity of the longevity associated transcriptional factor FOXO4, which regulates lifespan extension, tumor suppression, and energy metabolism. The encoded protein may also be a key regulator of blood pressure by controlling the transport of sodium and chloride ions. Specifically, the rs956868 SNP in the WNK1 gene is a missense variant, and has been shown to be associated with high blood pressure.

**NADSYN1 (rs2276362)**
The gene NADSYN1 regulates NAD+, an important cofactor that plays a central role in metabolism, best known for being a coenzyme in redox reactions and as a signaling molecule. NAD+ levels fall as we age, and the levels of NAD+ itself may be the reason for extended longevity. The NADSYN1 gene has been shown to regulate longevity in model organisms. The promoter of NADSYN1 has a FOXO3A recognition sequence and NADSYN1 transcription is promoted through phosphorylation of FOXO3A, one of the handful genes repeatedly shown to be associated with longevity. The rs2276362 SNP has been shown to be linked to higher vitamin D status in the human body, which in turn has been shown to promote protein homeostasis and hence also longevity.

In conclusions, the results of this study reveal new sides of the complex genetic regulation of longevity and suggest 47 new variants that could potentially be associated with longevity in the Bulgarian population as these are found in significantly higher allele frequency in the Bulgarian centenarian sample. Based on their assigned functional role, three of these genes were further investigated. These genes, RNF43, WNK1 and NADSYN1, are involved in evolutionary conserved processes with well ascertained association with longevity, i.e., Wnt signaling pathway, insulin/IGF-1 signal pathway and redox balancing processes.
respectively. Genes containing new longevity variants are linked to major subnets of longevity genes. This is only a pilot study, and these results will be augmented with data from whole genome sequencing of the same subjects that we plan to perform in the near future.

DECLARATIONS

Authors’ contributions
Made substantial contributions to conception and design of the study and performed data analysis and interpretation: Serbezov D, Balabanski L, Karachanak-Yankova S, Toncheva D
Performed data acquisition, as well as provided administrative, technical, and material support: Vazharova R, Nesheva D, Hammoudeh Z, Staneva R, Mihaylova M, Damyanova V, Antonova O, Nikolova D, Hadjidekova S

Availability of data and materials
Data used for the analyses in this article can be provided by the authors upon request.

Financial support and sponsorship
The Bulgarian centenarians project was funded by the National Science Fund of Bulgaria (DN 03/7/18.12.2016), and Bulgarian Ministry of Education and Science under the National Program for Research "Young Scientists and Postdoctoral Students".

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
The study was approved by the Ethics committee of the Medical University of Sofia and was found to be in accordance with the requirements of national and international legislation for conducting research involving human subjects, e.g., the Declaration of Helsinki. The subjects were also required to sign an informed consent before biological samples were taken.

Consent for publication
Not applicable.

Copyright
© The Author(s) 2020.

REFERENCES
1. Eatock D. Demographic outlook for the European Union 2019. European Union; 2019. Available from: https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_IDA(2019)637955. [Last accessed on 27 Sep 2020]
2. NSI. Statistical Reference Book. Bulgaria: National Statistical Institute; 2020. Available from: https://www.nsi.bg/en/content/18296%D0%BF%D1%83%D0%B1%D0%BB%D0%B8%D0%BA%D0%B0%D1%86%D0%B8%D1%8F/statistical-reference-book-2020-bulgarian-version. [Last accessed on 27 Sep 2020]
3. Herskind AM, McGue M, Holm NV, Sørensen TI, Harvald B, et al. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. Hum Genet 1996;97:319-23.
4. Deelen J, Beekman M, Holm NV, Sørensen TI, Harvald B, et al. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. Hum Genet 1996;97:319-23.
5. Sebasti P, Solovieff N, Dewan AT, Walsh KM, Puca A, et al. Genetic signatures of exceptional longevity in humans. PLoS One 2012;7:e29848.
6. Gikas I, Petratou D, Tavernarakis N. Longevity pathways and memory aging. Front Genet 2014;5:155.
7. Serbezov D, Balabanski L, Hadjidekova S, Toncheva D. Genomics of longevity: recent insights from research on centenarians. Biotechnol Biotechnol Equip 2018;32:159-66.
8. Santos-Lozano A, Sanchis-Gomar F, Pareja-Galeano H, Fiuza-Luces C, Emanuele E, et al. Where are supercentenarians located? A
worldwide demographic study. Rejuvenation Res 2015;18:14-9.

9. Schlotterer C, Tobler R, Kohler R, Nolte V. Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. Nat Rev Genet 2014;15:749-63.

10. Fracassetti M, Griffin PC, Willi Y. Validation of pooled whole-genome re-sequencing in arabidopsis lyrata. PLoS One 2015;10:e0140462.

11. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc 2015;10:1556-66.

12. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv 2019:531210.

13. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 1995;57:289-300.

14. Team RC. R: a language and environment for statistical computing Vienna 2018 Available from: https://www.r-project.org/

15. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, et al. The reactome pathway knowledgebase. Nucleic Acids Res 2009;37:W305-11.

16. Budovsky A, Craig T, Wang J, Tacutu R, Csordas A, et al. LongevityMap: a database of human genetic variants associated with longevity. Trends Genet 2013;29:59-60.

17. Chen J, Bardes EE, Aironow BJ, Jegga AG. ToppGene suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res 2009;37:W305-11.

18. Chen Y, Whetstone HC, Lin AC, Nadesan P, Wei Q, et al. Beta-catenin signaling plays a disparate role in different phases of fracture repair: implications for therapy to improve bone healing. PLoS Med 2007;4:e249.

19. Ito M, Yang Z, Andl T, Cui C, Kim N, et al. Wnt-dependent de novo hair follicle regeneration in adult mouse skin after wounding. Nature 2007;447:316-20.

20. Reya T, Clevers H. Wnt signalling in stem cells and cancer. Nature 2005;434:843-50.

21. Klotz LO, Sánchez-Ramos C, Prieto-Arroyo I, Urbánek P, Steinbrenner H, et al. Redox regulation of FoxO transcription factors. Redox Biol 2015;6:51-72.

22. Rasmussen MA, Egeberg A, Mikkelsen DT, Hovig E, et al. The demography of AluYb8 insertion variation. PLoS Genet 2015;11:e1005489.

23. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, et al. The reactome pathway knowledgebase. Nucleic Acids Res 2009;37:W305-11.

24. Budovsky A, Craig T, Wang J, Tacutu R, Csordas A, et al. LongevityMap: a database of human genetic variants associated with longevity. Trends Genet 2013;29:59-60.

25. Chen J, Bardes EE, Aironow BJ, Jegga AG. ToppGene suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res 2009;37:W305-11.

26. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, et al. The reactome pathway knowledgebase. Nucleic Acids Res 2009;37:W305-11.

27. Budovsky A, Craig T, Wang J, Tacutu R, Csordas A, et al. LongevityMap: a database of human genetic variants associated with longevity. Trends Genet 2013;29:59-60.

28. Chen J, Bardes EE, Aironow BJ, Jegga AG. ToppGene suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res 2009;37:W305-11.