Доклади на Българската академия на науките
Comptes rendus de l'Académie bulgare des Sciences
Tome 76, No 1, 2023

MEDICINE

Medical genetics

PRIORITIZATION OF GENETIC VARIANTS PREDISPOSING TO ALZHEIMER'S DISEASE IN YOUNG HEALTHY BULGARIAN INDIVIDUALS USING CENTENARIAN EXOMES

Dragomira Nikolova[#], Lubomir Balabanski, Dimitar Serbezov, Shima Mehrabian[*], Maria Petrova[*], Latchezar Traykov[*], Draga Toncheva

Received on July 28, 2020
Accepted on September 29, 2020

Abstract

The present study aims to identify pathogenic/risky variants in the exomes of healthy young individuals predisposing to Alzheimer's disease (AD) by their validation in centenarian's exomes. The group of Bulgarian centenarians included 32 individuals (mean age 102.4) without AD diagnosis and the matched control group of 61 young healthy individuals (mean age 21.9). Two DNA pools were constructed with equimolar amounts of DNA from each participant and whole exome sequencing analysis (WES) was performed. We examined the WES data for pathogenic or risky SNPs predisposing to AD. Of altogether 1929 variants selected from DisGeNet database for association with AD, only 174 (9.02%) were detected in our WES data. One hundred and fifty-two (152) SNPs were present in both studied groups, 8 in centenarians and 14 in controls only. Seven variants have significantly higher frequency in young individuals or absence in centenarians. Based on sufficiently unambiguous literature data, we nominated rs63750264 in *APP*, rs429358 in *APOE*, rs1800562 in *HFE* variants for predisposition to AD in young individuals. To a lesser extent, the carriership of rs2070045 in *SORL1*, rs6265 in *BDNF*, rs769449 in *APOE*, *rs4988514*

[#]Corresponding author.

This study is a part of a project number DN 03/7 from 18.12.2016 and KP-06-N33/5 from 13.12.2019, Bulgarian National Science Fund.
DOI:10.7546/CRABS.2023.01.11

in *SST* can increase the risk to AD development. These variants could find implication in the Alzheimer's disease estimation in young individuals before the onset of clinical symptoms.

Key words: whole exome sequencing (WES), Alzheimer's disease (AD), healthy individuals, variants, predisposition

Introduction. Alzheimer's disease (AD) is a multifactorial neurodegenerative disorder with a huge negative implication on society. On a worldwide scale more than 50 million people suffer from dementia and AD and only one in four people is diagnosed (2019 Alzheimer's statistics). The morphological characteristics are progressive neuron destruction, accumulation of beta amyloid plaques (Aβ) in the brain of patients, synaptic loss, hyperphosphorylated tau protein as neurofibrillary tangles (NFTs), as well as impaired glucose metabolism.

In the last ten years there were many research efforts to reveal the main players in the complex pathology of this disease, as well as to identify biomarkers for AD predisposition. Studies on AD patients revealed impaired expression of proteins as amyloid precursor protein (APP), apolipoprotein E (ApoE), presenilin, c-reactive protein, heat shock proteins (HSPs), etc., resulting in memory loss or neuron degradation [1]. As a polygenic disorder AD has a heterogeneous nature with participation of both environmental and genetic factors influencing the age of onset, severity of the disease, progression and outcome of the patients.

The known genetic factors predisposing to AD are genes with defective functions for amyloid precursor protein (APP), presenilin 1 (*PSEN1*) and presenilin 2 (*PSEN2*), found in large multi-generational families, as well as the apolipoprotein E (APOE) epsilon4 allele and the sortilin-related receptor (*SORL1*) gene [2]. Iron accumulation in multiple organs including the brain as well as the dysregulation of cholesterol homeostasis were also pointed as a possible cause for the neurodegeneration [2]. At least 21 additional genetic risk loci for the genetically complex form of AD were revealed by genome-wide association studies and massive parallel resequencing efforts. Pathogenic mutations in these genes are very rare and are responsible for a small fraction of cases of the disease [3] and many traits remain hidden because of the insufficient power of the methodology used [4]. Large meta-analyses made in recent years are also problematic as they rely on previous SNP panels of different quality. Until now conventional genetic tests are applied for detection of a faulty gene for familial Alzheimer's disease which is passed down in next generations. We assume that many young people over 18 years old without family history of Alzheimer can also carry risk genes and could have predisposition to Alzheimer's disease later in life [5].

In this study we apply WES on two groups without symptoms of Alzhaimer's disease: cohort of 61 young individuals (18–25 years old) and 32 centenarians (100–106 years old) in order to identify the carriership of known pathogenic mutations predisposing to AD development. We hypothesized that the group of

healthy younger individuals may have genetic factors predisposing to Alzheimer's disease later in their life which can be absent among the centenarians' group. It is well known that centenarian exomes are a "gold standard" for healthy life and it is largely accepted that they do not contain pathogenic variants for chronic or devastating serious illnesses in their lifetime.

Materials and methods. The study received Ethics committee approval of the Medical University-Sofia, Bulgaria. Participants declared their willingness to participate in the study by signing a written informed consent. Interviews were conducted to receive additional information for their healthy status and medical history. Blood or buccal swab samples were collected from 93 healthy Bulgarian subjects divided into two groups according to their age: 32 centenarians (mean age 102.4) and 61 healthy individuals (mean age 21.9) referred to as "controls". Whole genome DNAs have been extracted and two pools constructed by using equimolar amounts of DNA from each individual.

The WES was conducted as previously described [6]. The total number of variants annotated in both pools after applying these filters was 89 810 (72 791 in both pools, 8253 in centenarian pool only and 8766 in control pool only).

DisGeNET database is an online database which contains information of human gene-disease associations (GDAs) as well as variant-disease associations (VDAs) from various resources including monogenic, polygenic and environmental diseases (`https://www.disgenet.org`). It contains 72 870 variant-disease associations (VDAs), between 46 589 SNPs and 6356 diseases and phenotypes. We selected altogether 1929 variants studied for association with Alzheimer's disease from DisGenNET. The presence and frequency of those variants were inspected in the WES data of Bulgarian healthy subjects.

Results. Altogether 174 variants (in 127 genes) out of 1929 AD variants were found in Bulgarian WES data, of them 152 variants in both pools, 8 variants only in centenarians and 14 variants – only in healthy adults aged 18–30 years. We used Fisher's exact test to determine the significance of allele frequency differences between variants present in the two pools.

The frequencies of 152 variants were compared between the two studied groups. The correlation was quite high ($r = 0.88$) which indicates that these variants probably have a small effect on the disease development since they are present in the genomes of centenarians without AD diagnosis (Fig. 1). Only few variants are located at a significant distance from the main cluster of variants distributed by allele frequencies.

The significance of differences in the frequencies of variants between the two pools of Bulgarian centenarians and controls is visualized by Manhattan plot (Fig. 2).

Variants which showed statistically significant higher frequency in controls compared to centenarians were further analyzed for association with AD on the base of literature data. We selected four missense variants described to be AD risk

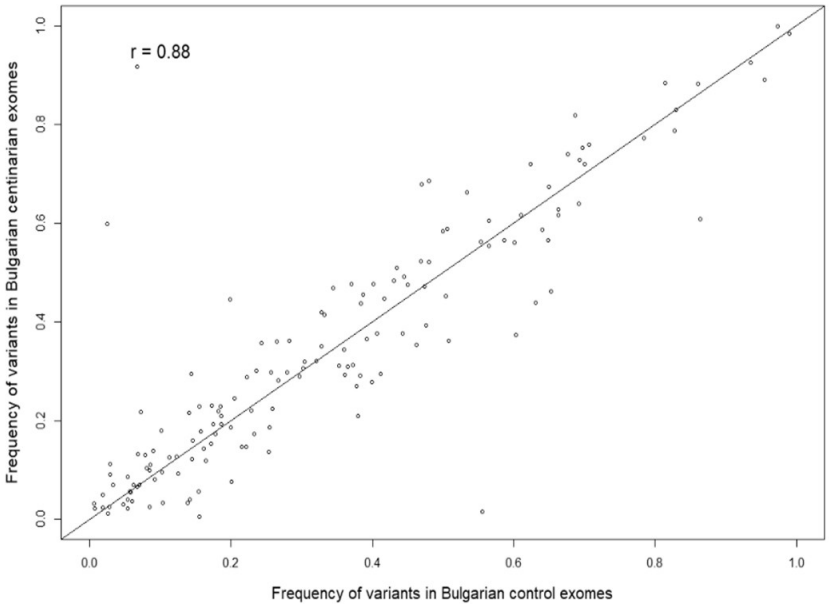

Fig. 1. 152 variants associated with AD divided by allele frequency between the centenarians and controls (correlation coefficient $r = 0.88$)

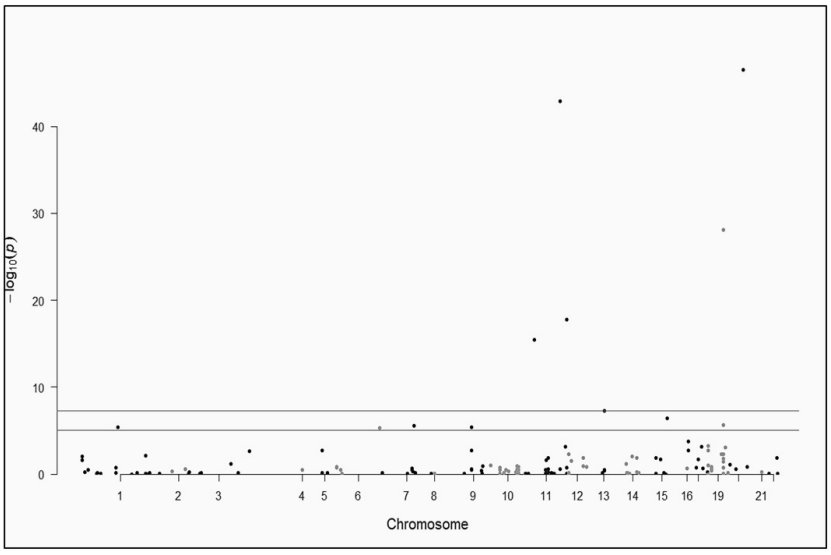

Fig. 2. Manhattan plot of 152 variants associated with AD called in both centenarian and control pools. The horizontal lines correspond to $p = 1.0 \times 10^{-5}$ and 5.0×10^{-8}

factors. Variants rs2070045 in *SORL1* and rs6265 in *BDNF* have been determined as risk factors for AD by many other research groups (Table 1).

Table 1

Variants selected to be predisposed to Alzheimer's disease in our study. First four variants in genes *SORL1, BDNF, APOE* and *SST* exhibit a significant difference between centenarians and young healthy subjects; the last three variants in *APOE, APP, HFE* are pathogenic or risk factors for AD absent in the group of centenarians

CHR	GENE	REF	ALT	LOCATION	RS	MAF in centenarians (> 100 years)	MAF in young healthy adults (18–30 years)	FDR
11	*SORL1*	T	G	exonic	rs2070045	0.005	0.155	$1.6801.E^{-18}$
11	*BDNF*	C	T	exonic	rs6265	0.076	0.201	$3.4178.E^{-16}$
19	*APOE*	G	A	intronic	rs769449	0.033	0.104	$2.4543.E^{-06}$
3	*SST*	T	C	UTR3	rs4988514	0.041	0.142	$2.4859.\ E^{-03}$
CHR	GENE	REF	ALT	LOCATION	RS	MAF in centenarians (> 100 years)	MAF in young healthy adults (18–30 years)	FDR
19	*APOE*	T	C	exonic	rs429358	–	0.056	–
21	*APP*	C	T	exonic	rs63750264	–	0.003	–
6	*HFE*	G	A	exonic	rs1800562	–	0.026	–

Abbreviation: Chr. – chromosome, Ref – referent allele, Alt – alternative allele, MAF – minor allele frequency, FDR-false discovery rate adjusted p-value

The rest 148 variants are not considered pathogenic/risk factors associated with the disease as they have either similar frequencies in both groups or a higher frequency in the group of centenarians.

Among variants detected only in young healthy subjects (14 variants) two variants were pathogenic (*APP* rs63750264, *APOE* rs429358) and one (*HFE* rs1800562) – risk factor for AD in ClinVar databases. Their clinical significance was validated in a large number of publications (Table 1).

In the group of centenarians only we detected 8 variants: rs193922916 (*APP*), rs3219012 (*OGG1*), rs35988749 (*NTF3*), rs5030732 (*UCHL1*), rs533451404 and rs63751294 (*GRN*), rs58973334 (*PSEN2*), rs778084883 (*PAX5*). There is inconsistent or controversial publicly available data about their clinical significance. The carriership of those variants in non-AD centenarians suggests that they are not independent risk factors for AD development.

Discussion. The common way of unveiling genetic factors playing a significant role in AD is to study individuals who have already developed the disorder and/or families with significant cases of the disease running in their members such as GWAS used cases-controls comparison. Recently NGS technology helped to identify hundreds of variants with potential impact on the disease risk and progression. Those approaches resulted in a generation of a vast majority of genes and variants of known or still not fully clarified clinical significance. The strategy of our study was to validate genetic variants with evidence from previous research studies for association with AD using centenarians without clinical AD diagnosis as a tool. The aim of the study was to detect carriership of possibly pathogenic and/or risk variants which will be helpful for early diagnosis for AD in the future. One thousand nine hundred and twenty-nine variants suggested as AD associated by DicGeNET database were checked for their presence and frequency in both groups of healthy individuals. We accepted the carriership of seven variants to be important as risk conferring factors in young individuals.

Variants detected in both groups. Allele G of *SORL1* rs2070045 is considered a risk factor for AD and MCI (mild cognitive impairment) as it influences the expression of APP and endothelial lipoprotein in different regions of the brain. The risk G-allele carriers have a decreased positive rsFC (resting-state functional connectivity) between the hippocampus and middle temporal gyrus compared with TT carriers [7]. A larger meta-analysis from 2016 [8] reported rs2070045*G to be connected with an increased risk of AD. Therefore, *SORL1* rs2070045 is associated with cognitive ageing but its effects could be different in both sexes.

The variant rs6265 in *BDNF* has been studied in relation to bipolar disorders, depressiveness and Parkinson's disease. Recent data connects it to AD pathogenesis, suggesting a synergistic action between BDNF (rs6265) and DBH (rs1611115) contributing to a greater AD risk [9]. However, we could not confirm such a trait as the significance of the variant in our study is detected in an exome pool of individuals independently of their sex.

Variants rs769449 in *APOE* and rs4988514 in *SST* showed lower statistical difference between centenarians and controls and there is less evidence for their involvement in AD (Table 1). Nevertheless, rs769449 in *APOE* was one of the three intronic variants that reached exome-wide significance in a recent study of late-onset AD [10]. Similarly, rs4988514 in SST has been reported to be associated with sporadic AD and a risk modifier in two distant world populations [11, 12]. That is why we consider it important to report those two variants (rs769449 in *APOE*, rs4988514 in *SST*) as risk factors possibly contributing to AD in our population. We presume that the variants detected in both groups most probably work synergistically to define higher risk in healthy individuals without expressed clinical symptoms of AD.

Variants absent in centenarians. rs429358 (T/C) is a functional polymorphism leading to a missense mutation (p. Cys130Arg) in *APOE*. An investigation between cholesterol metabolism and brain functional alterations in individuals with AD identified rs429358 as a core genetic variation associated with disease-related differences in the brain function [13]. Nowadays rs429358 is accepted as a "top genetic risk factor" for AD [14]. For the first time the SNP was declared as an independent risk factor for LOAD (due to the weak disequilibrium with rs7412) and significantly associated with the tau levels in the cerebrospinal fluid [15].

rs63750264 in *APP* gene represents a rare variant site where all three possible single nucleotide mutations are known and considered pathogenic (c.2149G>T, p. Val717Phe or V717F; c.2149G>A, p. Val717Ile or V716I; and c.2149G>C, p. Val717Leu or V717L). The mutation in exon 17 of *APP* gene causes abnormal APP processing with subsequent overproduction of amyloid beta protein. Therefore, APP is directly involved in the pathogenesis of AD. The link between APP and APOE was also studied suggesting that APOE promotes both the deposition and fibrillization of amyloid beta affecting clearance of protease-resistant Abeta/apoE aggregates [16].

Variant rs1800562 (p.Cys282Tyr) in exon 4 of *HFE* gene was described as a pathogenic variant for hemochromatosis [17]. *HFE* C282Y allele leads to an excess of redox-active iron and the induction of oxidative stress in neurons which is aggravated in carriers of apolipoprotein E epsilon4 allele. The authors stated the synergistic action between HFE C282Y and C2 allele of transferrin gene. Thus each of those three variants is supposedly considered a risk factor for AD [18]. In 2012 rs1800562 in *HFE* was reported as a susceptibility AD locus in ADNI (AD Neuroimaging Initiative) study of 400 patients with mild cognitive impairment and 200 early onset AD patients. The association between iron metabolism, the level of the oxidative stress in the brain and AD was mentioned again in 2014, as MRI studies revealed increased accumulation of iron in carriers of HFE mutations [19]. Iron is particularly dense in the areas close to the beta-amyloid plaques and mutant NFE protein is unable to limit the uptake of iron in the brain [19]. The expression of C282Y-HFE alters the normal cell homeostasis by increasing the iron

uptake in the brain, therefore, the level of the oxidative stress, cell proliferation and cholesterol. All these findings suggest that HFE is directly connected to the iron level and indirectly to AD in the population.

In conclusion, in our study we examine exomes of healthy individuals (centenariancs and young persons) to determine the clinical significance of variants that are defined/suspected to be associated with Alzheimer's disease in publicly available literature sources. Of 1929 from DisGeNET database 174 variants were found in our exome data. Their prioritization for predisposition to AD is based on their significantly higher frequency in young individuals or absence in centenarians, as well as detailed literature survey. Thus seven variants are considered significant disease-predisposing factors. They could be used as genetic markers for presymptomatic early determination of AD in young healthy individuals at risk.

REFERENCES

[1] Kumar D., A. Sharma, L. Sharma (2020) A comprehensive review of Alzheimer's association with related proteins: Pathological role and therapeutic significance, Curr. Neuropharmacol., **18**(8), 674–695.

[2] Reitz C., R. Mayeux (2009) Endophenotypes in normal brain morphology and Alzheimer's disease: a review, Neuroscience, **164**(1), 174–190.

[3] Reitz C., R. Mayeux (2010) Use of genetic variation as biomarkers for mild cognitive impairment and progression of mild cognitive impairment to dementia, J. Alzheimers Dis., **19**(1), 229–251.

[4] Medway C., H. Shi, J. Bullock, H. Black, K. Brown et al. (2010) Using In silico LD clumping and meta-analysis of genome-wide datasets as a complementary tool to investigate and validate new candidate biomarkers in Alzheimer's disease, Int. J. Mol. Epidemiol. Genet., **1**(2), 134–144.

[5] Ganev M., D. Serbezov, S. Karachanak-Yankova, R. Vazharova, D. Nesheva et al. (2019) Prioritization of genetic variants predisposing to coronary heart disease in the Bulgarian population using centenarian exomes, Biotechnology & Biotechnological Equipment, **33**(1), 1757–1765.

[6] Balabanski L., D. Serbezov, D. Nikolova, O. Antonova, D. Nesheva et al. (2020) Centenarian Exomes as a Tool for Evaluating the Clinical Relevance of Germline Tumor Suppressor Mutations, Technol. Cancer Res. Treat., doi: 10.1177/1533033820911082.

[7] Shen J., W. Qin, Q. Xu, L. Xu, J. Xu et al. (2017) Modulation of APOE and SORL1 genes on hippocampal functional connectivity in healthy young adults, Brain Struct. Funct., **222**(6), 2877–2889.

[8] Wang Z., H. Lei, M. Zheng, Y. Li, Y. Cui et al. (2016) Meta-analysis of the Association between Alzheimer Disease and Variants in GAB2, PICALM, and SORL1, Mol. Neurobiol., **53**(9), 6501–6510.

[9] Belbin O., K. Morgan, C. Medway, D. Warden, M. Cortina-Borja et al. (2019) The Epistasis Project: A Multi-Cohort Study of the Effects of BDNF, DBH, and SORT1 Epistasis on Alzheimer's Disease Risk, J. Alzheimers Dis., **68**(4), 1535–1547.

[10] Jakobsdottir J., S. J. van der Lee, J. C. Bis, V. Chouraki, D. Li-Kroeger et al. (2016) Rare Functional Variant in TM2D3 is Associated with Late-Onset Alzheimer's Disease, PLoS Genet., **12**(10), e1006327.
[11] Xue S., L. Jia, J. Jia (2009) Association between somatostatin gene polymorphisms and sporadic Alzheimer's disease in Chinese population, Neurosci. Lett., **465**(2), 181–183.
[12] Vepsalainen S., S. Helisalmi, A. M. Koivisto, T. Tapaninen, M. Hiltunen et al. (2007) Somatostatin genetic variants modify the risk for Alzheimer's disease among Finnish patients, J. Neurol., **254**(11), 1504–1508.
[13] Bai F., Y. Yuan, Y. Shi, Z. Zhang (2016) Multiple genetic imaging study of the association between cholesterol metabolism and brain functional alterations in individuals with risk factors for Alzheimer's disease, Oncotarget., **7**(13), 15315–15328.
[14] Hao X., J. Yan, X. Yao, S. L. Risacher, A. J. Saykin et al. (2016) Diagnosis-Guided Method for Identifying Multi-Modality Neuroimaging Biomarkers Associated with Genetic Risk Factors in Alzheimer's Disease, Pac. Symp. Biocomput., **21**, 108–119.
[15] Rao S., M. Ghani, Z. Guo, Y. Deming, K. Wang et al. (2018) An APOE-independent cis-eSNP on chromosome 19q13.32 influences tau levels and late-onset Alzheimer's disease risk, Neurobiol. Aging, **66**, 178 e1–178 e8.
[16] Bales K. R., T. Verina, D. J. Cummins, Y. Du, R. C. Dodel et al. (1999) Apolipoprotein E is essential for amyloid deposition in the APP(V717F) transgenic mouse model of Alzheimer's disease, Proc. Natl. Acad. Sci. USA, **96**(26), 15233–15238.
[17] Barton J. C., C. Q. Edwards (1993) HFE Hemochromatosis. In: GeneReviews (eds M. P. Adam et al.), Seattle (WA).
[18] Robson K. J., D. J. Lehmann, V. L. Wimhurst, K. J. Livesey, M. Combrinck et al. (2004) Synergy between the C2 allele of transferrin and the C282Y allele of the haemochromatosis gene (HFE) as risk factors for developing Alzheimer's disease, J. Med. Genet., **41**(4), 261–265.
[19] Ali-Rahmani F., C. L. Schengrund, J. R. Connor (2014) HFE gene variants, iron, and lipids: a novel connection in Alzheimer's disease, Front. Pharmacol., doi: 10.3389/fphar.2014.00165.

Department of Medical Genetics
Medical Faculty
Medical University – Sofia
15 Akad. Ivan Geshov Blvd
Sofia, Bulgaria
e-mail: dmb@abv.bg
lubobal@yahoo.co.uk
dimitase@gmail.com
dragatoncheva@gmail.com

*Department of Neurology
University Hospital "Alexandrovska"
1 Georgi Sofiiski St
1431 Sofia, Bulgaria
e-mail: shima_meh@yahoo.com
dr.mpetrova@yahoo.com
traykov_l@yahoo.fr