SHORT COMMUNICATION

Variability and positive selection in FOXP2, a gene associated with the development of language, speech, and cognition

Variabilidad y selección positiva en FOXP2, gen asociado al desarrollo del lenguaje, el habla y la cognición

Sergio V. Flores[1,2] ✉, Alicia Figueroa-Barra[2,3] ✉, María Labraña-Palma[4] ✉, Angel Roco-Videla[5] ✉, Marcela Caviedes-Olmos[6] ✉, Sofía Perez-Jiménez[6] ✉, Raúl Aguilera Eguía[7] ✉

[1]Universidad Arturo Prat. Santiago, Chile.

[2]Universidad de Chile, Departamento de Psiquiatría y Salud Mental, Laboratorio de Psiquiatría Traslacional (Psiquislab) Universidad de Chile. Santiago, Chile.

[3]Núcleo Milenio para Mejorar la Salud Mental de Adolescentes y Jóvenes, Imhay. Santiago, Chile.

[4]Universidad Autónoma de Chile, Facultad de Ciencias de la Salud. Santiago, Chile.

[5]Universidad Bernardo O´Higgins, Programa de Magister en Ciencias Químico-Biológicas. Santiago, Chile.

[6]Universidad de las Américas. Facultad de Salud y Ciencias Sociales, Santiago, Chile.

[7]Universidad Católica de la Santísima Concepción, Departamento de Salud Pública, Concepción, Chile.

ABSTRACT

Introduction: the FOXP2 gene has been identified as a key genetic factor in the development of language and human cognition. Mutations in FOXP2 have been associated with language disorders and speech difficulties. Additionally, this gene has been linked to various neuropsychiatric conditions. The objective of this study is to analyze the genetic differentiation of populations in the FOXP2 gene and in the rs10447760, rs1456031, rs2253478 and rs2396753 polymorphisms.
Method: data from the "1000 Genomes" Project were used to analyze genetic variability in FOXP2 in 2504 individuals from 26 populations and 5 macro populations. Linkage disequilibrium, Hardy-Weinberg equilibrium and allele frequencies of the SNPs were evaluated. Genetic differentiation was estimated using the FST statistic.
Results: a highly differentiated region was identified in intron 3 of FOXP2 between the African macro population and the rest, with a maximum FST of 0,78. This region contains an epigenetic mark H3K27Ac, suggesting a regulatory role. Hardy-Weinberg imbalances were observed in some populations for the SNPs analyzed. Linkage disequilibrium analysis showed that these SNPs have independent effects.
Conclusions: the highly differentiated region in FOXP2 suggests a past natural selection event, supporting an adaptive role of this gene in the evolution of language, speech and cognition. Population differences in Hardy-Weinberg equilibrium and genetic variability highlight the importance of considering genetic variation in future association studies with FOXP2.

Keywords: Genetic Differentiation; FOXP2; Language Evolution; SNP Analysis; Positive Selection.

RESUMEN

Introducción: el gen FOXP2 se ha identificado como un factor genético clave en el desarrollo del lenguaje y la cognición humana. Mutaciones en FOXP2 se han asociado con trastornos del lenguaje y dificultades del habla. Además, este gen se ha vinculado a diversas condiciones neuropsiquiátricas. El objetivo de este estudio es analizar la diferenciación genética de poblaciones en el gen FOXP2 y en los polimorfismos rs10447760,

rs1456031, rs2253478 y rs2396753.

Método: se utilizaron datos del Proyecto "1000 Genomas" para analizar la variabilidad genética en FOXP2 en 2504 individuos de 26 poblaciones y 5 macro poblaciones. Se evaluó el desequilibrio de ligamiento, el equilibrio de Hardy-Weinberg y las frecuencias alélicas de los SNPs. La diferenciación genética se estimó mediante el estadístico FST.

Resultados: se identificó una región altamente diferenciada en el intrón 3 de FOXP2 entre la macro población africana y el resto, con un FST máximo de 0,78. Esta región contiene una marca epigenética H3K27Ac, sugiriendo un papel regulatorio. Se observaron desequilibrios de Hardy-Weinberg en algunas poblaciones para los SNPs analizados. El análisis de desequilibrio de ligamiento mostró que estos SNPs tienen efectos independientes.

Conclusiones: la región altamente diferenciada en FOXP2 sugiere un evento de selección natural pasado, apoyando un papel adaptativo de este gen en la evolución del lenguaje, el habla y la cognición. Las diferencias poblacionales en el equilibrio de Hardy-Weinberg y la variabilidad genética resaltan la importancia de considerar la variación genética en futuros estudios de asociación con FOXP2.

Palabras clave: Diferenciación Genética; FOXP2; Evolución del Lenguaje; Análisis de SNP; Selección Positiva.

INTRODUCTION

The FOXP2 (Forkhead Box P2) gene has been identified as a genetic factor in the development of human language and cognition. FOXP2 regulates genes involved in the formation and function of the nervous system, particularly in brain areas related to language and fine motor skills.[1] Mutations in FOXP2 have been associated with language disorders and speech difficulties.[2] Beyond its role in language development, FOXP2 has also been linked to various neuropsychiatric conditions.[3,4]

Recent studies have identified several genetic markers in FOXP2 associated with neuropsychiatric conditions and cognitive deficits.[5,6] The rs10447760 polymorphism, for example, is located in the 5' regulatory region of FOXP2 and has shown associations with linguistic and cognitive functions, particularly in patients with chronic schizophrenia[7], as well as differences in cognitive performance and clinical outcomes in schizophrenia patients.[8] Additionally, this genetic marker may interact with body mass index (BMI) to influence cognitive deficits observed in these patients.[7,9]

Other SNPs, such as rs1456031, rs2253478, and rs2396753, have also been investigated for their potential impact on cognitive function and the development of language disorders. These polymorphisms seem to contribute to the expression of linguistic and cognitive phenotypes.[3,10] Variations in these SNPs are associated with differences in susceptibility to neuropsychiatric disorders, suggesting a modulatory role of these genetic variants in brain development and function.[4,11] The presence of pleiotropy in determining cognitive abilities and various mental health disorders indicates that these SNPs can influence multiple phenotypic aspects.[2,12]

The aim of this study is to analyze the genetic differentiation of populations in the FOXP2 gene, as well as in the SNPs rs10447760, rs1456031, rs2253478, and rs2396753. Understanding the genetic variability in these markers is important to comprehend how genetic differences can influence language and cognitive-related phenotypes in different populations, as well as to expand our knowledge on the evolution of language and provide evidence for future genetic association studies.

METHOD

For this study, data from the "1000 Genomes Project," one of the most genomically dense publicly available databases, were used. This database includes information from 2504 individuals distributed across 26 populations and 5 macro populations: Africa, America, East Asia, South Asia, and Europe. It contains more than 84 million SNPs (Single Nucleotide Polymorphisms), used for genetic analysis. The complete FOXP2 gene, spanning 607,445 nucleotides and 25731 SNPs, was downloaded. Table 1 details the sample used in this study.

Tabla1. Samples analyzed in this study		
Macro Population	**Population**	**Sample Size**
Africa (AFR)	Esan in Nigeria (ESN)	99
	Gambian in Western Division, The Gambia (GWD)	113
	Luhya in Webuye, Kenya (LWK)	99
	Mende in Sierra Leone (MSL)	85
	Yoruba in Ibadan, Nigeria (YRI)	108

	African Ancestry in Southwest US (ASW)	61
	African Caribbean in Barbados (ACB)	96
	African Caribbean in Barbados (ACB)	114
	Total AFR	775
East Asia (EAS)	Southern Han Chinese (CHS)	105
	Kinh in Ho Chi Minh City, Vietnam (KHV)	99
	Japanese in Tokyo, Japan (JPT)	104
	Han Chinese South (CHS)	105
	Han Chinese in Beijing, China (CHB)	103
	Total EAS	516
South Asia (SAS)	Bengali in Bangladesh (BEB)	86
	Sri Lankan Tamil in the UK (STU)	96
	Indian Telugu in the UK (ITU)	102
	Gujarati Indians in Houston, Texas (GIH)	103
	Punjabi in Lahore, Pakistan (PJL)	96
	Total SAS	483
Europe (EUR)	Utah Residents (CEPH) with Northern and Western European Ancestry (CEU)	99
	Toscani in Italia (TSI)	107
	Iberian Population in Spain (IBS)	107
	British in England and Scotland (GBR)	91
	Finnish in Finland (FIN)	99
	Total EUR	1105
Latin America (AMR)	Puerto Rican in Puerto Rico (PUR)	104
	Colombian in Medellin, Colombia (CLM)	94
	Mexican Ancestry in Los Angeles, California (MXL)	64
	Peruvian in Lima, Peru (PEL)	85
	Total AMR	411
	Grand Total	2504

Linkage disequilibrium between the four SNPs in this study was evaluated through a correlation analysis, providing information on genetic recombination. Additionally, Hardy-Weinberg equilibrium was analyzed both globally and within each macro population and individual population. This analysis is important to identify potential deviations caused by evolutionary factors such as selection, mutation, migration, population structuring, or genetic drift. The risk allele frequencies for each SNP were calculated to gain detailed insight into how these alleles are distributed in different population contexts.

Genetic differentiation was estimated using the Weir and Cockerham`s FST estimator across the entire FOXP2 gene. This analysis was performed with sliding windows of 2,000 bp to identify regions within the gene with significant variability among populations, which may suggest regions subject to natural selection in the past.

All analyses mentioned were performed using the VCFtools program and the statistical software R. VCFtools was used to manipulate and analyze VCF (Variant Call Format) files, while R was used for statistical analysis and data visualization.

RESULTS

The estimation of FST in sliding windows identified a region of high differentiation between the "Africa" macro population and the rest of the macro populations (figure 1). This region is located in intron 3 of FOXP2, spanning 23 kb and containing 644 SNPs. The highest FST was 0,78, found in the window ranging from position 114030000 to 114032000, which includes two SNPs: rs1818998 and rs1818999. There is no evidence of association between these markers and phenotypes in the literature. For the total SNPs of FOXP2 analyzed here, the median FST was 0,1690, 95% CI [0,1749, 0,1871], whereas for the region of high differentiation, these values were much higher: 0,5653, 95% CI [0,5058, 0,6012]. The Mann-Whitney U test to compare the medians of the two data sets indicated a very significant difference between both sets (P= $1,58 \times 10^{-24}$). When analyzing the region of high differentiation in the UCSC Genome Browser, an H3K27Ac epigenetic mark was found at the center of this region.
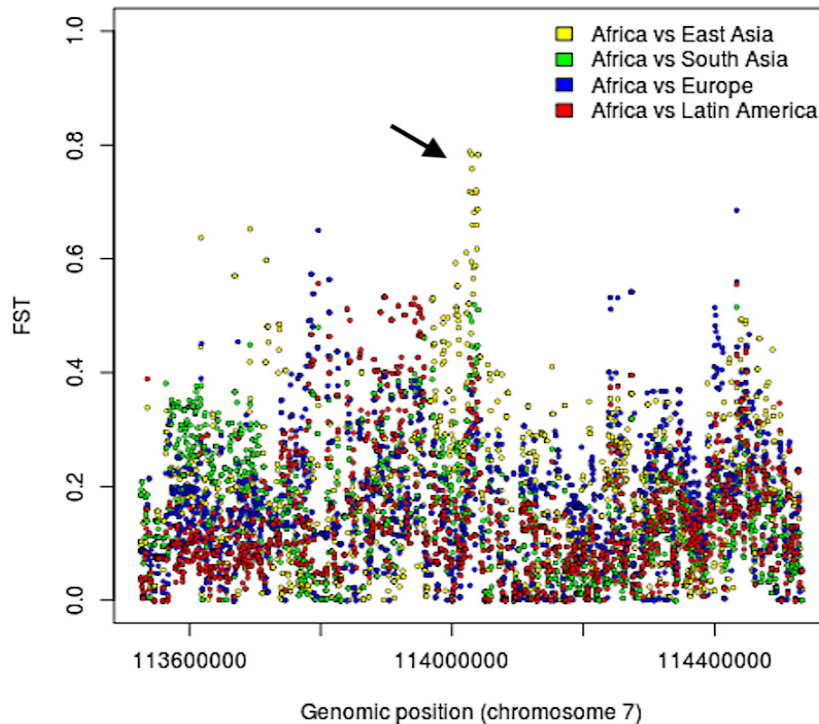
Figure 1. Estimation of FST across the FOXP2 gene and 200 kb flanking the gene

The Hardy-Weinberg equilibrium analysis for SNP rs10447760 showed significant global disequilibrium (P_{Xi}^2 = 7,98 x 10^{-10}). SNP rs1456031 also presented significant global disequilibrium (P_{Xi}^2 = 2,39 x 10^{-5}), while only the AFR macro population showed disequilibrium (P_{Xi}^2 = 0,006), as well as Chinese Dai in Xishuangbanna (EAS) at the population level (P_{Xi}^2 = 0,005). For SNP rs2253478, significant global disequilibrium was also observed (P_{Xi}^2 = 1,63 x 10^{-6}), with no cases identified in macro populations, but in the Gujarati Indian in Houston (SAS) population (P_{Xi}^2 = 0,021). SNP rs2396753 showed disequilibrium only at the global level (P_{Xi}^2 = 0,00198), whereas for the two highly differentiated SNPs identified in this study, disequilibrium was observed only at the global level (P_{Xi}^2 = 9,93 x 10^{-34}). Figure 2 shows the allele frequencies of these SNPs.

Finally, the linkage disequilibrium analysis determined that the four SNPs are in equilibrium. The highest correlation was found in European populations, with r^2= 0,27 between SNPs rs2253478 and rs2396753, which are separated by 170335 bp. Among the two highly differentiated SNPs reported in this study (rs1818998 and rs1818999), a strong correlation (r^2= 0,998) was found, being only 230 bp apart. The ancestral and major haplotypes for both SNPs are rs1818998 T/rs1818999 T and rs1818998 C/rs1818999 C. Only two individuals, both from South Asia, were found to carry a recombinant chromosome (rs1818998 C/rs1818999 T): one from Bengali in Bangladesh and the other from Gujarati Indians in Houston.

DISCUSSION

The highly differentiated region identified in intron 3 of the FOXP2 gene suggests a past natural selection event. This region, spanning 23 kb and containing 644 SNPs, shows the highest FST of 0,78 in the window ranging from position 114030000 to 114032000, indicating significant genetic differentiation. The identification of an H3K27Ac epigenetic mark in the center of this region supports the hypothesis that it may be involved in the regulation of gene expression. This epigenetic mark was not empirically detected in previous studies[13,14] is associated with the activation of gene expression and is frequently found in promoter and enhancer regions, which are important for regulating gene transcription. This pattern of differentiation is important as it suggests an adaptive role for FOXP2.

Regarding Hardy-Weinberg equilibrium, the observed disequilibria in AFR, Gujarati Indian in Houston, Chinese Dai in Xishuangbanna, and Gujarati Indian in Houston populations may reflect local variations in population dynamics and could be due to genetic structuring resulting in non-random mating patterns, possibly influenced by historical and geographical factors.[15] The heterogeneity of allele frequencies (figure 2) indicates that it is relevant to consider genetic variation in future studies with these markers, as well as to explore the effect of genetic ancestry on them.
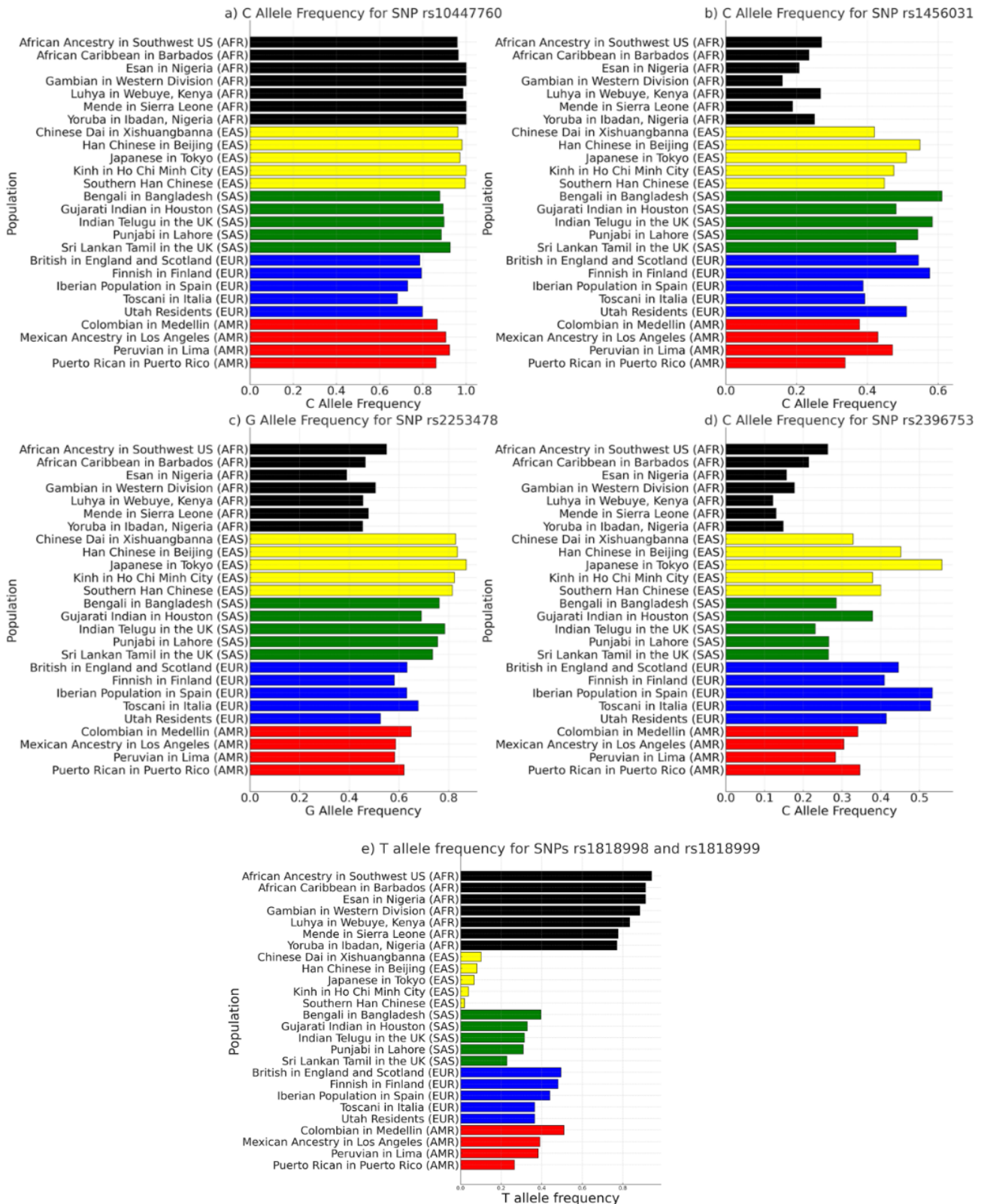
Figure 2. Allele frequencies in different populations for SNPs in the FOXP2 gene

In the linkage disequilibrium (LD) analysis, the results showed that, in general, these SNPs are in equilibrium in most populations, suggesting that they are not co-segregating, thus having independent effects on the phenotype. Among the two highly differentiated SNPs identified in this study (rs1818998 and rs1818999), an almost total correlation was found, which is explained by their proximity.

CONCLUSIONS

The highly differentiated region in intron 3 of the FOXP2 gene in the African population suggests a past natural selection event, supporting the hypothesis of an adaptive role in the evolution of language, speech, and cognition.

The deviation from Hardy-Weinberg equilibrium in certain populations reflects differences in population structure and possible historical and geographical influences, indicating the importance of considering genetic variation in future studies.

The genetic variability of the markers studied here shows significant differences between macro populations, which is important for future association studies with FOXP2.

REFERENCES

1. Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature. 2001;413(6855):519-23. https://doi.org/10.1038/35097076.

2. Hagenaars SP, Harris SE, Davies G, Hill WD, Liewald DC, Ritchie SJ, et al. Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112,151) and 24 GWAS consortia. Mol Psychiatry. 2016;21(11):1624-32. https://doi.org/10.1038/mp.2015.225.

3. McCarthy-Jones S, Green MJ, Scott RJ, Tooney PA, Cairns MJ, Wu JQ, et al. Preliminary evidence of an interaction between the FOXP2 gene and childhood emotional abuse predicting likelihood of auditory verbal hallucinations in schizophrenia. J Psychiatr Res. 2014;50:66-72. https://doi.org/10.1016/j.jpsychires.2013.11.012.

4. Mozzi A, Forni D, Clerici M, Pozzoli U, Mascheretti S, Guerini FR, et al. The evolutionary history of genes involved in spoken and written language: beyond FOXP2. Sci Rep. 2016;6:22157. https://doi.org/10.1038/srep22157.

5. Misiak B, Stramecki F, Gawęda Ł, Prochwicz K, Sąsiadek MM, Moustafa AA, et al. Interactions between variation in candidate genes and environmental factors in the etiology of schizophrenia and bipolar disorder: a systematic review. Mol Neurobiol. 2018;55(6):5075-5100. https://doi.org/10.1007/s12035-017-0708-y.

6. Mueller KL, Murray JC, Michaelson JJ, Christiansen MH, Reilly S, Tomblin JB. Common genetic variants in FOXP2 are not associated with individual differences in language development. PLoS One. 2016;11(4). https://doi.org/10.1371/journal.pone.0152576.

7. Yang M, Cui Y, Xue M, Forster MT, Lang X, Xiu M, et al. Sexual dimorphism in the relationship between Forkhead-Box P2 and BMI with cognitive deficits in schizophrenia. Front Aging Neurosci. 2022;14:920352. https://doi.org/10.3389/fnagi.2022.920352.

8. Lang X, Zhang W, Song X, Zhang G, Du X, Zhou Y, et al. FOXP2 contributes to the cognitive impairment in chronic patients with schizophrenia. Aging (Albany NY). 2019;11(16):6440-6448. https://doi.org/10.18632/aging.102198.

9. Li T, Zeng Z, Zhao Q, Wang T, Huang K, Li J, et al. FoxP2 is significantly associated with schizophrenia and major depression in the Chinese Han population. World J Biol Psychiatry. 2013;14(2):146-150. https://doi.org/10.3109/15622975.2011.615860.

10. Španiel F, Horáček J, Tintěra J, Ibrahim I, Novák T, Čermák J, et al. Genetic variation in FOXP2 alters grey matter concentrations in schizophrenia patients. Neurosci Lett. 2011;493(2):131-135. https://doi.org/10.1016/j.neulet.2011.02.024.

11. Tolosa A, Sanjuán J, Dagnall AM, Moltó MD, Herrero N, de Frutos R. FOXP2 gene and language impairment in schizophrenia: association and epigenetic studies. BMC Med Genet. 2010;11:114. https://doi.org/10.1186/1471-2350-11-114.

12. Sanjuán J, Tolosa A, González JC, Aguilar EJ, Pérez-Tur J, Nájera C, et al. Association between FOXP2 polymorphisms and schizophrenia with auditory hallucinations. Psychiatr Genet. 2006;16(2):67-72. https://doi.org/10.1097/01.ypg.0000185029.35558.bb.

13. Becker M, Devanna P, Fisher SE, Vernes SC. Mapping of Human FOXP2 Enhancers Reveals Complex

Regulation. Front Mol Neurosci. 2018;21(11):47. doi: https://10.3389/fnmol.2018.00047.

14. Torres-Ruiz R, Benítez-Burraco A, Martínez-Lage M, Rodríguez-Perales S, García-Bellido P. Functional characterization of two enhancers located downstream FOXP2. BMC Med Genet. 2019;20(1):65. https://10.1186/s12881-019-0810-2

15. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526: 68–74 (2015). https://doi.org/10.1038/nature15393