Arch Med Sadowej Kryminol 2022; vol. 73 (2): 61-66 DOI: 10.4467/16891716AMSIK.22.008.16806

archiwum medycyny sądowej i kryminologii

Praca oryginalna
Original paper

Katarzyna Linkowska[1], Boris A. Malyarchuk[2], Miroslava V. Derenko[2], Tomasz Grzybowski[1]

An association between copy number variation of enhancer involved in craniofacial development and biogeographic ancestry

Związek pomiędzy zmiennością liczby kopii wzmacniacza zaangażowanego w rozwój twarzoczaszki a pochodzeniem biogeograficznym

1. Department of Forensic Medicine, Division of Molecular & Forensic Genetics, Faculty of Medicine, Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Toruń, Poland
2. Institute of Biological Problems of the North, Far-East Branch of the Russian Academy of Sciences, Magadan, Russia

Abstract

Human facial morphology is a combination of many complex traits and is determined by a large number of genes and enhancers. Here, we report a Copy Number Variation (CNV) study of enhancer hs1431 in populations of Central European and South Siberian ancestry. Central European samples included 97 Poles, while South Siberian samples included 78 Buryats and 27 Tuvinians. CNVs were detected by real-time PCR, using ViiA™ 7 Real-Time PCR System (Applied Biosystems). We revealed significant differences in CNV of hs1431 enhancer between Polish and Buryat population (p=0.0378), but not between Central European and South Siberian population (p=0.1225). Our results suggest that an increase in copy number variation of hs1431 enhancer is associated with biogeographic ancestry. However, this result needs extending and replicating in larger cohorts. This is the first study revealing the presence of copy number variation of enhancer hs1431 in humans.

Key words: CNV; enhancer; facial shape; biogeographic ancestry; craniofacial development

Nadesłano: 2022-08-29
Zaakceptowano: 2022-10-20

Streszczenie

Morfologia ludzkiej twarzy jest połączeniem wielu złożonych cech i determinowana jest przez dużą liczbę genów i wzmacniaczy. W niniejszym badaniu analizowano zmienność liczby kopii (CNV) wzmacniacza hs1431 w populacjach pochodzenia środkowoeuropejskiego i południowo-syberyjskiego. Próbki z Europy Środkowej pochodziły od 97 Polaków, natomiast próbki z południowej Syberii pochodziły od 78 Buriatów i 27 Tuwańczyków. CNV analizowano metodą real-time PCR z wykorzystaniem aparatu ViiA™ 7 (Applied Biosystems). Statystycznie istotne różnice w liczbie kopii wzmacniacza hs1431 występowały między populacjami polską a buriacką (p=0,0378), ale nie między populacjami Europy Środkowej i południowej Syberii (p=0,1225). Uzyskane wyniki sugerują, że wzrost liczby kopii wzmacniacza hs1431 jest związany z pochodzeniem biogeograficznym. Jednak wynik ten wymaga rozszerzenia i powtórzenia na większych kohortach. Jest to pierwsze badanie, w którym wykazano zmienności liczby kopii wzmacniacza hs1431 u ludzi.

Słowa kluczowe: CNV; wzmacniacz; kształt twarzy; pochodzenie biogeograficzne; rozwój twarzoczaszki

1. Introduction

Prediction of externally visible characteristics based on genetic data is challenging for forensic genetics. In contrast to pigmentation traits, which can often be predicted with high accuracy, determining the shape of the face is still under study. Highly complex nature of facial morphology is determined by joint action of a large number of coding and non-coding elements of the genome. Both mutations affecting the protein coding sequence and copy number changes are likely to influence abnormal morphology of the face [1-4]. However, genetic basis influencing normal facial shape is still poorly understood. Accumulating evidence from genome-wide association studies (GWAS) indicates that sequence variation in non-coding regions strongly contributes to facial shape [5,6]. The study of genetic architecture of facial shape variation in mice identified that non-pathological changes in craniofacial morphology are caused by deletion of craniofacial enhancers [7]. Distant-acting mammalian enhancers are non-coding regulatory elements important for the temporal and spatial in vivo expression of genes. They can be located tens to hundreds of thousands of base pairs away from their target genes [8].

Attanasio et al. (2013) suggested that variation in the sequence or copy number of craniofacial enhancers may contribute to the spectrum of facial variation observed in human populations. Copy Number Variations (CNVs) are known to be important contributors to phenotypic diversity. CNVs have a higher per-locus mutation rate than SNPs (Single Nucleotide Polymorphisms) and are likely responsible for a significant amount of both human gene and genome evolution, as well as genetic diversity between individuals—perhaps to a larger extent than SNPs [9]. Moreover, a decrease or an increase in copy number may cause different phenotypic outcomes [2,9].

A better understanding of genetic basis of facial shape will improve our knowledge of the complex relationship between genotype and phenotype in craniofacial abnormalities and provide a basis for predicting facial features in numerous applications, including medical diagnosis, aesthetic medicine and forensics. The aim of our study was to determine the CNV of enhancer hs1431 in Central European and South Siberian population samples to ascertain whether its deletion or duplication may be related to their biogeographic ancestry. We selected hs1431 enhancer located in regulatory vicinity of negative transcriptional regulator SNAI2 involved in craniofacial development because its deletion resulted in the most severe reduction in craniofacial gene expression (Snai2) in Attanasio et al. (2013) study and there are no additional known enhancers with overlapping activity. Deletion of hs1431 in mice producing phenotypic effects involves multiple regions of the skull (e.g. an increase in facial length and in the width of anterior neurocranium, and a shortening of anterior cranial base) but much less severe than observed upon deletion of the Snai2 itself [7]. SNAI2 belongs to the Snail family of zinc finger transcription factors and plays an important role in developmental processes. The post-natal expression of SNAI2 and the effects of its deletion and overexpression are similar in mice and humans [10].

2. Materials and Methods

2.1. Study samples

Our study included 202 DNA samples from unrelated individuals of Central European and South Siberian ancestry. Central European samples included 97 Poles, while South Siberian samples included 78 Buryats and 27 Tuvinians.

Hair samples from 27 Tuvinians were collected in Dzun-Khemchiksk, Mongun-Taiga, Bai-Taiga, Ovyursk, Tes-Khemsk, Erzinsk and Tandinsk districts of Tuva Republic. Most of the territory of the Tuva Republic is situated in the steppe zone in the centre of the Asian continent, bounded by the Sayan Mountains to the North and the Mongolian steppes to the South.

Buryats, who are also nomadic cattle breeders, live in the central southern part of the Siberia bordering Mongolia and China. Blood samples from 78 Buryats were collected in different localities of the Buryat Republic (Dzhida, Bichursk, Ivolga, Kyakhta, Tunka, Kizhinga, Khorinsk, Zakamensk, Eravna, Selenga, Barguzin and Kabansk districts), Chita (Aginsk district) and Irkutsk (Alarsk, Bayandaevsk, Bokhan, Nukutsk, Olkhon, Ust-Orda, Osa and Ekhirit-Balagansk districts) regions, thus encompassing all territories inhabited by the modern Buryats.

All individuals were paternally and maternally unrelated through at least three generations and originated from the areas considered for this study. These samples were previously analyzed for mtDNA and Y-chromosome variation [11-14].

The study was approved by the Bioethics Committee of Nicolaus Copernicus University (consent number KB 504/2014). All methods were performed in accordance with the relevant guidelines and regulations. Informed consent was obtained from all individual participants included in the study.

2.2. Genetic analysis

CNV analysis of the hs1431 enhancer region was performed by real-time PCR, using F and R primers (F: 5'-TGCCAGTTCAGAGAACCATTAGAC-3'; R: 5'- GGAGCGGTTCCTGGTGAAG-3'), 6-FAM dye-labeled probe (Probe 5'-CTTCCTGGA-ATCTTT-3') and ViiA™ 7 Real-Time PCR System (Applied Biosystems, Foster City, California, USA), according to the TaqMan® Copy Number Assays Protocol. TaqMan Copy Number Assay was run simultaneously with TaqMan Copy Number Reference Assay human RNase P (Applied Biosystems) in a duplex reaction. The reference assay detects the Ribonuclease P RNA component H1 (H1RNA) gene (RPPH1) on chromosome 14, which is known to exist in two copies in diploid genome. The PCR reaction contained 10 ng of template DNA, 5 µl TaqMan® Genotyping Master Mix (Applied Biosystems), 0.5 µl TaqMan® Copy Number Reference Assay, 500 nM concentration of primers F and R, 250 nM concentration of 6-FAM dye-labeled probe in a final volume of 10 µl. The amplification conditions were: 10 min at 95°C and 40 cycles of 95°C for 15 sec and 60°C for 1 min. Post-PCR data analysis was performed by using CopyCaller Software (Applied Biosystems) to determine the number of copies of the target sequence in each genomic DNA sample.

2.3. Statistical analysis

The Fisher exact test was used to compare both populations. This test is applied in analysis of small samples and it is necessary when one or more expected values are less than 5. The significance level for statistical tests was 0.05. The statistical calculations were performed using Statistica package v.12.5 (Statsoft).

3. Results

The analysis of CNV revealed that the variant consisting of two copies of hs1431 enhancer occurs with the highest frequency in all samples (Table 1).

Table 1. CNV variant frequencies of hs1431 enhancer

CNV variants	European population	Asian population		p-value
	Poles (N=97)	Buryats (N=78)	Tuvinians (N=27)	
2 copies	1.000	0.949	1.000	0.1225
3 copies	0.000	0.051	0.000	

Three copies of hs1431 enhancer were observed only in 4 samples from the Buryat population. No deletions of hs1431 enhancer were detected. Statistical analysis showed significant differences in copy number variation of hs1431 enhancer between

Polish and Buryat populations (p=0.0378). However, no statistically important differences were found in CNV of hs1431 between Central European (Poles) and South Siberian population (Buryats and Tuvinians considered as one sample, p=0.1225) and between Tuvinian and Buryat population (p=0.5703).

4. Discussion

Numerous studies have shown associations between genetic factors and normal facial shape variation. These include candidate gene studies focusing on a small number of genetic loci with known roles in facial development or anomalies [15] and genome-wide association studies (GWASs) that examined millions of genetic polymorphisms [6]. Several approaches to phenotyping of facial shape features were used in association studies. There is currently no consensus on the optimal strategy in that regard. The variety of different methods used to describe facial shape makes it difficult to compare results and may partly explain the lack of replication across different studies [16]. An innovative, data-driven facial phenotyping approach based on structural correlations between about 10,000 3D quasi-landmarks, which enabled the hierarchical (global-to-local) clustering of the human face into segments was proposed by Claes at al. The proposed methods are based on the indirect prediction of facial phenotypes, with ancestry and sex prediction DNA data playing a key role in this regard [17,18]. Taking into account the lack of a consistent approach to describing facial phenotypic features and knowing that the ancestry informative markers that are associated with different phenotypic features, show differences in the variant frequencies at the continental level [19], we focused our preliminary research on analyzing CNV frequency in Central European and South Siberian populations.

Claes et al. (2018) demonstrated an association of facial variation and regulatory elements, especially distal enhancers. They suggested that the genetic variation within an overlapping set of regulatory elements may influence both species-specific and individual facial shape variation in humans [18]. Differences in copy number of enhancer hs1431 between Polish and Buryat population actually imply that hs1431 enhancer may play a role in facial morphology in humans. However, due to lack of statistically significant differences between Polish and South Siberian populations, we cannot exclude the presence of additional unknown enhancers with overlapping activity. It is known that enhancer redundancy is a remarkably widespread feature of mammalian genomes that provides an effective regulatory buffer to prevent deleterious phenotypic consequences. The pervasive presence of multiple enhancers with similar activities near the same gene confers phenotypic robustness to loss-of-function mutations in individual enhancers, therefore, they will cause mostly subtle phenotypes in humans [20]. Statistically significant differences between Polish and Buryat populations and lack of significant differences between Polish and Tuvinian populations may possibly be attributed to the known diversity of the Siberian gene pool. Indeed, several population studies employing different genetic markers revealed that populations of southwest Siberia (including Tuvinians) are marked by the highest influx of West Eurasian (European) genetic component, while in southeast Siberians (including Buryats) this contribution was markedly lower [13,19,21]. In this respect, it is worth noting that early physical anthropological studies on the populations of the Altaian and Sayan regions of Siberia, unambiguously suggested a directional but not uniform replacement of the "Europeoid" component by the "Mongoloid" one. In the ancestors of present-day Buryats "Mongoloid"(East Asian) features had been prevalent since Neolithic [13]. Finally, a recent ancestry informative markers (AIMs) study revealed that the structure of Siberian population results from many factors including ancient migration patterns, cultural distinctiveness, small population sizes and low population densities [19].

The main limitation of our study is small sample size. One cannot exclude that the lack of significant differences between Polish and Tuvinian populations as well as between Polish and South Siberian populations may be due to the sample size which is not great enough to get a statistically significant result. Indeed, our analysis had only 0.49 power to detect the differences between CNVs frequencies of hs1431. In order to prove statistical significance, at least 202 individuals from each population should be tested. Therefore, this result needs to be extended

in larger cohorts of a defined biogeographic ancestry.

The major value of the study lies in detection of copy variation of enhancer hs1431. There are multiple scenarios for such copy number variation, for example, duplication of the enhancer in close proximity of the original gene, duplication of the PCR-target sequence within or outside of the enhancer or whole locus duplication including the gene. Not all of the possibilities translate into straightforward changes of gene expression. Thus, further studies are needed to determine the nature of this variation.

Lack of differences in copy number of enhancer hs1431 between Central European and South Siberian population samples does not exclude its role in contributing to the normal facial shape variation. Extended sample size as well as inclusion of other enhancers and samples from additional populations will be needed for a better understanding of the genetics of human facial morphology and functional significance of regulatory complexity of enhancers. Ultimately, more detailed knowledge in that area might contribute to future forensic applications. Indeed, the genetic prediction of physical appearance traits occupies an important place in contemporary forensic research. Today, forensic science can benefit from the rapidly developing methods in the areas of genomics and machine learning, which is particularly beneficial for further development of genetic prediction of physical traits. Moreover, the evolution of the approach, which has moved from building predictive models based on variables that show genetic association to building models based on variables, allows better prediction of physical appearance traits from DNA [22].

References

1. Dixon, M.J.; Marazita, M.L.; Beaty, T.H.; Murray, J.C. Cleft lip and palate: understanding genetic and environmental influences. *Nat Rev Genet* **2011**, *12*, 167-178, doi:10.1038/nrg2933.
2. Hammond, P.; McKee, S.; Suttie, M.; Allanson, J.; Cobben, J.M.; Maas, S.M.; Quarrell, O.; Smith, A.C.; Lewis, S.; Tassabehji, M.; et al. Opposite effects on facial morphology due to gene dosage sensitivity. *Hum Genet* **2014**, *133*, 1117-1125, doi:10.1007/s00439-014-1455-z.
3. Ng, S.B.; Buckingham, K.J.; Lee, C.; Bigham, A.W.; Tabor, H.K.; Dent, K.M.; Huff, C.D.; Shannon, P.T.; Jabs, E.W.; Nickerson, D.A.; et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* **2010**, *42*, 30-35, doi:10.1038/ng.499.
4. Satokata, I.; Maas, R. Msx1 deficient mice exhibit cleft palate and abnormalities of craniofacial and tooth development. *Nat Genet* **1994**, *6*, 348-356, doi:10.1038/ng0494-348.
5. Qiao, L.; Yang, Y.; Fu, P.; Hu, S.; Zhou, H.; Peng, S.; Tan, J.; Lu, Y.; Lou, H.; Lu, D.; et al. Genome-wide variants of Eurasian facial shape differentiation and a prospective model of DNA based face prediction. *J Genet Genomics* **2018**, *45*, 419-432, doi:10.1016/j.jgg.2018.07.009.
6. Shaffer, J.R.; Orlova, E.; Lee, M.K.; Leslie, E.J.; Raffensperger, Z.D.; Heike, C.L.; Cunningham, M.L.; Hecht, J.T.; Kau, C.H.; Nidey, N.L.; et al. Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology. *PLoS Genet* **2016**, *12*, e1006149, doi:10.1371/journal.pgen.1006149.
7. Attanasio, C.; Nord, A.S.; Zhu, Y.; Blow, M.J.; Li, Z.; Liberton, D.K.; Morrison, H.; Plajzer-Frick, I.; Holt, A.; Hosseini, R.; et al. Fine tuning of craniofacial morphology by distant-acting enhancers. *Science* **2013**, *342*, 1241006, doi:10.1126/science.1241006.
8. Dickel, D.E.; Visel, A.; Pennacchio, L.A. Functional anatomy of distant-acting mammalian enhancers. *Philos Trans R Soc Lond B Biol Sci* **2013**, *368*, 20120359, doi:10.1098/rstb.2012.0359.
9. Stankiewicz, P.; Lupski, J.R. Structural variation in the human genome and its role in disease. *Annu Rev Med* **2010**, *61*, 437-455, doi:10.1146/annurev-med-100708-204735.
10. Perez-Mancera, P.A.; Bermejo-Rodriguez, C.; Gonzalez-Herrero, I.; Herranz, M.; Flores, T.; Jimenez, R.; Sanchez-Garcia, I. Adipose tissue mass is modulated by SLUG (SNAI2). *Hum Mol Genet* **2007**, *16*, 2972-2986, doi:10.1093/hmg/ddm278.
11. Derenko, M.; Denisova, G.; Malyarchuk, B.; Dambueva, I.; Bazarov, B. Mitogenomic diversity and differentiation of the Buryats. *J Hum Genet* **2018**, *63*, 71-81, doi:10.1038/s10038-017-0370-2.
12. Derenko, M.; Malyarchuk, B.; Denisova, G.A.; Wozniak, M.; Dambueva, I.; Dorzhu, C.; Luzina, F.; Miscicka-Sliwka, D.; Zakharov, I. Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai-Sayan regions. *Hum Genet* **2006**, *118*, 591-604, doi:10.1007/s00439-005-0076-y.
13. Derenko, M.V.; Grzybowski, T.; Malyarchuk, B.A.; Dambueva, I.K.; Denisova, G.A.; Czarny, J.; Dorzhu, C.M.; Kakpakov, V.T.; Miscicka-Sliwka, D.; Wozniak, M.; et al. Diversity of mitochondrial DNA lineages in South Siberia. *Ann Hum Genet* **2003**, *67*, 391-411, doi:10.1046/j.1469-1809.2003.00035.x.

14. Wozniak, M.; Derenko, M.; Malyarchuk, B.; Dambueva, I.; Grzybowski, T.; Miscicka-Sliwka, D. Allelic and haplotypic frequencies at 11 Y-STR loci in Buryats from South-East Siberia. *Forensic Sci Int* **2006**, *164*, 271-275, doi:10.1016/j.forsciint.2005.11.023.
15. Boehringer, S.; van der Lijn, F.; Liu, F.; Gunther, M.; Sinigerova, S.; Nowak, S.; Ludwig, K.U.; Herberz, R.; Klein, S.; Hofman, A.; et al. Genetic determination of human facial morphology: links between cleft-lips and normal variation. *Eur J Hum Genet* **2011**, *19*, 1192-1197, doi:10.1038/ejhg.2011.110.
16. Lee, M.K.; Shaffer, J.R.; Leslie, E.J.; Orlova, E.; Carlson, J.C.; Feingold, E.; Marazita, M.L.; Weinberg, S.M. Genome-wide association study of facial morphology reveals novel associations with FREM1 and PARK2. *PLoS One* **2017**, *12*, e0176566, doi:10.1371/journal.pone.0176566.
17. Claes, P.; Liberton, D.K.; Daniels, K.; Rosana, K.M.; Quillen, E.E.; Pearson, L.N.; McEvoy, B.; Bauchet, M.; Zaidi, A.A.; Yao, W.; et al. Modeling 3D facial shape from DNA. *PLoS Genet* **2014**, *10*, e1004224, doi:10.1371/journal.pgen.1004224.
18. Claes, P.; Roosenboom, J.; White, J.D.; Swigut, T.; Sero, D.; Li, J.; Lee, M.K.; Zaidi, A.; Mattern, B.C.; Liebowitz, C.; et al. Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nat Genet* **2018**, *50*, 414-423, doi:10.1038/s41588-018-0057-4.
19. Rogalla, U.; Rychlicka, E.; Derenko, M.V.; Malyarchuk, B.A.; Grzybowski, T. Simple and cost-effective 14-loci SNP assay designed for differentiation of European, East Asian and African samples. *Forensic Sci Int Genet* **2015**, *14*, 42-49, doi:10.1016/j.fsigen.2014.09.009.
20. Osterwalder, M.; Barozzi, I.; Tissieres, V.; Fukuda-Yuzawa, Y.; Mannion, B.J.; Afzal, S.Y.; Lee, E.A.; Zhu, Y.; Plajzer-Frick, I.; Pickle, C.S.; et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **2018**, *554*, 239-243, doi:10.1038/nature25461.
21. Derenko, M.V.; Malyarchuk, B.A.; Wozniak, M.; Dambuveva, I.K.; Dorzhu, C.M.; Luzina, F.A.; Lee, H.K.; Miscicka-Sliwka, D.; Zakharov, I.A. The diversity of Y-chromosome lineages in indigenous population of South Siberia. *Dokl Biol Sci* **2006**, *411*, 466-470, doi:10.1134/s0012496606060123.
22. Pośpiech, E.; Teisseyre, P.; Mielniczuk, J.; Branicki, W. Predicting Physical Appearance from DNA Data-Towards Genomic Solutions. *Genes (Basel)* **2022**, *13*, doi:10.3390/genes13010121.

ORCID

Katarzyna Linkowska – 0000-0001-7632-4181
Boris A. Malyarchuk – 0000-0002-0304-0652
Miroslava V. Derenko – 0000-0002-1849-784X
Tomasz Grzybowski – 0000-0001-6228-6460

CORRESPONDING AUTHOR

Katarzyna Linkowska
Department of Forensic Medicine, Division of Molecular & Forensic Genetics, Collegium Medicum in Bydgoszcz
Curie-Sklodowskiej Str. 9, Bydgoszcz 85-094, Poland
tel. +48 5853552
linkowska@cm.umk.pl