Gerontology

Experimental Section / Original Paper

Gerontology
DOI: 10.1159/000509471

Received: February 16, 2020
Accepted: June 17, 2020
Published online: September 2, 2020

Natural Selection at the NHLH2 Core Promoter Exceptionally Long CA-Repeat in Human and Disease-Only Genotypes in Late-Onset Neurocognitive Disorder

Hossein Afshar[a] Fatemeh Adelirad[b] Ali Kowsari[c] Naser Kalhor[c] Ahmad Delbari[a] Reza Najafipour[d] Mahshid Foroughan[a] Ali Bozorgmehr[e] Safoura Khamse[a] Neda Nazaripanah[b] Mina Ohadi[a]

[a]Iranian Research Center on Aging, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran; [b]Department of Health Education and Promotion, Faculty of Health Sciences Tabriz University of Medical Sciences, Tabriz, Iran; [c]Department of Mesenchymal Stem Cell, The Academic Center for Education, Culture and Research, Qom, Iran; [d]Cellular and Molecular Research Centre, Research Institute for Prevention of Non Communicable Disease, Qazvin University of Medical Sciences, Qazvin, Iran; [e]Department of Neuroscience, Faculty of Advanced Technologies in Medicine, Iran University of Medical Sciences, Tehran, Iran

Keywords
NHLH2 · Short tandem repeat · Neurocognitive disorder · Natural selection

Abstract
Background: Approximately 2% of the human core promoter short tandem repeats (STRs) reach lengths of ≥6 repeats, which may in part be a result of adaptive evolutionary processes and natural selection. A single-exon transcript of the human nescient helix loop helix 2 (*NHLH2*) gene is flanked by the longest CA-repeat detected in a human protein-coding gene core promoter (Ensembl transcript ID: ENST00000369506.1). NHLH2 is involved in several biological and pathological pathways, such as motivated exercise, obesity, and diabetes. ***Methods:*** The allele and genotype distribution of the *NHLH2* CA-repeat were investigated by sequencing in 655 Iranian subjects, consisting of late-onset neurocognitive disorder (NCD) as a clinical entity (n = 290) and matched controls (n = 365). The evolutionary trend of the CA-repeat was also studied across vertebrates. ***Results:*** The allele range was between 9 and 25 repeats in the NCD cases, and 12 and 24 repeats in the controls. At the frequency of 0.56, the 21-repeat allele was the predominant allele in the controls. While the 21-repeat was also the predominant allele in the NCD patients, we detected significant decline of the frequency ($p < 0.0001$) and homozygosity ($p < 0.006$) of this allele in this group. Furthermore, 12 genotypes were detected across 16 patients (5.5% of the entire NCD sample) and not in the controls (disease-only genotypes; $p < 0.0003$), consisting of at least one extreme allele. The extreme alleles were at 9, 12, 13, 18, and 19 repeats (extreme short end), and 23, 24, and 25 repeats (extreme long end), and their frequencies ranged between 0.001 and 0.04. The frequency of the 21-repeat allele significantly dropped to 0.09 in the disease-only genotype compartment ($p <$

karger@karger.com
www.karger.com/ger

Mina Ohadi
Iranian Research Center on Aging
University of Social Welfare and Rehabilitation Sciences
Tehran 1985713834 (Iran)
mi.ohadi@uswr.ac.ir

0.0001). Evolutionarily, while the maximum length of the *NHLH2* CA-repeat was 11 repeats in non-primates, this CA-repeat was ≥14 repeats in primates and reached maximum length in human. ***Conclusion:*** We propose a novel locus for late-onset NCD at the *NHLH2* core promoter exceptionally long CA-STR and natural selection at this locus. Furthermore, there was indication of genotypes at this locus that unambiguously linked to late-onset NCD. This is the first instance of natural selection in favor of a predominantly abundant STR allele in human and its differential distribution in late-onset NCD.

Introduction

In comparison to single nucleotide substitutions, short tandem repeats (STRs) offer a significantly more versatile reservoir of genetic variations that may be necessary for speciation and species-specific phenotypes [1, 2]. We have previously reported that approximately 2% of the human protein-coding genes contain STRs of ≥6 repeats in the critical core promoter region [3]. Preliminary findings indicate that expansion of a number of these STRs may be linked to adaptive evolutionary processes in human, and also endure the burden of certain human disorders such as schizophrenia, bipolar disorder, and Alzheimer's disease [4–7].

At the top of the list of the genes consisting of exceptionally long STRs, the nescient helix loop helix 2 gene (*NHLH2*) contains a CA-repeat of 24 repeats in its core promoter, at position –4 from the transcription start site (TSS), which flanks a single-exon transcript of the gene (Ensembl transcript ID: ENST00000369506.1; Fig. 1). The *NHLH2* CA-repeat, along with the CA-repeat in the *SCGB2B2* gene, is the longest human CA-STR identified in a human protein-coding gene core promoter [8].

NHLH2 is predominantly expressed in hypothalamus, and is involved in several biological processes, a number of which are considered as protective or risk factors for the pathogenesis of various types of late-onset neurocognitive disorders (NCDs), including Alzheimer's disease, Lewy body dementia, vascular dementia, and frontotemporal dementia (Fig. 2) [9–16].

Here, we studied the distribution of the *NHLH2* gene core promoter CA-repeat in a group of patients with late-onset NCD and matched controls.

Materials and Methods

Subjects

Six hundred fifty-five unrelated Iranian subjects, consisting of late-onset NCD patients (n = 290) and controls (n = 365), were recruited from the provinces of Qazvin and Rasht. All patients were included based on the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). In each case, the Persian version of the Abbreviated Mental Test Score (AMTS) [17, 18] was implemented (AMTS of ≤7 was an inclusion criterion for NCD), medical records were reviewed in all participants, and CT scans were taken where possible (40% of instances). The AMTS is currently one of the most accurate primary screening instruments to increase the probability of NCD [19]. The Persian version of the AMTS is a valid cognitive assessment tool for older Iranian adults and can be used for NCD screening in Iran [18]. The control group was selected based on normal cognitive AMTS of >7, lack of major medical history, and normal CT scan where possible. The cases and controls were matched based on age and residential district. The mean age of the NCD cases and controls was 75.5 and 78.4, respectively. The subjects' consent was obtained (from their guardians where necessary) and their identities remained confidential throughout the study. The research was approved by the responsible Ethical Committee of Social Welfare and Rehabilitation Sciences, Tehran, Iran, and was consistent with the principles outlined in an internationally recognized standard for the ethical conduct of human research.

Allele and Genotype Analysis of the NHLH2 Gene CA-Repeat

Genomic DNA was obtained from peripheral blood using a standard salting out method. The following primers were used to amplify the region containing the *NHLH2* CA-STR:

Forward: TGCTCATTGTAGTTTGGCCAC

Reverse: CTCATGTTCAGTAAGTGTTGGCT

PCR reactions were carried out in a thermocycler (Creacon, model 0005.401) under the following conditions: 95 °C for 3 min,

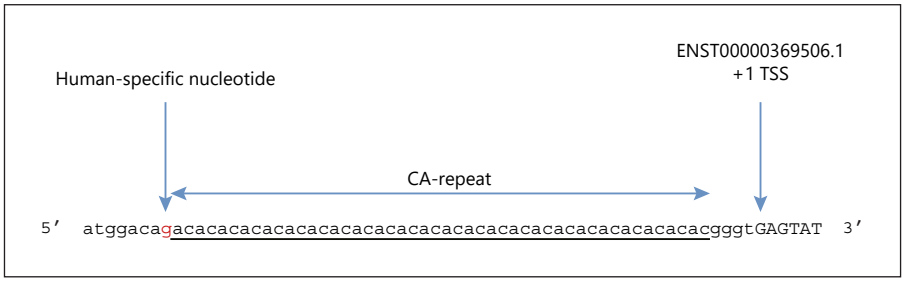

Fig. 1. Schematic representation of the *NHLH2* exceptionally long CA-repeat based on the Ensembl database. The CA-repeat 3′ end is at position –5 from the +1 TSS.