



OPEN

## Natural selection at the *RASGEF1C* (GGC) repeat in human and divergent genotypes in late-onset neurocognitive disorder

Z. Jafarian<sup>1</sup>, S. Khamse<sup>1</sup>, H. Afshar<sup>1✉</sup>, H.R. Khorram Khorshid<sup>2</sup>, A. Delbari<sup>1</sup> & M. Ohadi<sup>1✉</sup>

Expression dysregulation of the neuron-specific gene, *RASGEF1C* (RasGEF Domain Family Member 1C), occurs in late-onset neurocognitive disorders (NCDs), such as Alzheimer's disease. This gene contains a (GGC)<sub>13</sub>, spanning its core promoter and 5' untranslated region (*RASGEF1C*-201 ENST00000361132.9). Here we sequenced the (GGC)-repeat in a sample of human subjects (N = 269), consisting of late-onset NCDs (N = 115) and controls (N = 154). We also studied the status of this STR across various primate and non-primate species based on Ensembl 103. The 6-repeat allele was the predominant allele in the controls (frequency = 0.85) and NCD patients (frequency = 0.78). The NCD genotype compartment consisted of an excess of genotypes that lacked the 6-repeat (divergent genotypes) (Mid-P exact = 0.004). A number of those genotypes were not detected in the control group (Mid-P exact = 0.007). The *RASGEF1C* (GGC)-repeat expanded beyond 2-repeats specifically in primates, and was at maximum length in human. We conclude that there is natural selection for the 6-repeat allele of the *RASGEF1C* (GGC)-repeat in human, and significant divergence from that allele in late-onset NCDs. STR alleles that are predominantly abundant and genotypes that deviate from those alleles are underappreciated features, which may have deep evolutionary and pathological consequences.

### Abbreviations

AD	Alzheimer's disease
AMTS	Abbreviated Mental Test Score
MMSE	Mini-Mental State Exam
NCD	Neurocognitive disorder
<i>RASGEF1C</i>	RasGEF Domain Family Member 1C
STR	Short tandem repeat
TSS	Transcription start site
UTR	Untranslated region

Short tandem repeats (STRs), also known as microsatellites/simple sequence repeats, are an important source of evolutionary and pathological processes<sup>1–5</sup>. Numerous STRs within gene regulatory regions may have a link with the evolution of human and non-human primates through various mechanisms, such as gene expression regulation<sup>6–9</sup>. In a number of instances, there are indications of a link between STRs and late-onset neurocognitive disorders (NCDs) in human such as Alzheimer's disease (AD) and Parkinson's disease (PD)<sup>10–12</sup>.

*RASGEF1C* (RasGEF Domain Family Member 1C), located on chromosome 5q35.3, contains a (GGC)-repeat of 13-repeats, spanning its core promoter and 5' UTR (*RASGEF1C*-201 ENST00000361132.9)<sup>13</sup>. Based on Ensembl 103 (ensembl.org), the transcript containing the (GGC)-repeat is at the highest support level annotated for the transcript isoforms of this gene (TSL:1). The protein encoded by *RASGEF1C* is a guanine nucleotide exchange factor (GEF) (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=RASGEF1C>), and

<sup>1</sup>Iranian Research Center on Aging, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. <sup>2</sup>Personalized Medicine and Genometabolomics Research Center, Hope Generation Foundation, Tehran, Iran. ✉email: drhoseinafshar@yahoo.com; mi.ohadi@uswr.ac.ir

primarily interacts with a number of the ZNF family members, such as ZNF507, ZNF235, ZNF25, and ZNF612 (<https://version11.string-db.org/cgi/network.pl?taskId=QkiV955revRw>). In human, *RASGEF1C* is predominantly expressed in the brain (<https://www.proteinatlas.org/ENSG00000146090-RASGEF1C/tissue>), and aberrant regulation of this gene occurs in late-onset NCDs, such as AD<sup>14</sup>.

Abundance of polymorphic CGG repeats in the human genome suggest a broad involvement in neurological diseases<sup>15</sup>. Large GGC expansions are strictly linked to neurological disorders of predominant neurocognitive impairment, such as fragile X-associated tremor and ataxia syndrome, NIID, and oculopharyngeal muscular dystrophy<sup>16–21</sup>.

Here we sequenced the *RASGEF1C* (GGC)-repeat in a sample of humans, consisting of late-onset NCDs and controls. We also analyzed the status of this STR across several primate and non-primate species.

## Materials and methods

**Subjects.** Two hundred sixty-nine unrelated Iranian subjects of  $\geq 60$  years of age, consisting of late-onset NCD patients ( $n = 115$ ) and controls ( $n = 154$ ) were recruited from the provinces of Tehran, Qazvin, and Rasht. In each NCD case, the Persian version<sup>22</sup> of the Abbreviated Mental Test Score (AMTS)<sup>23</sup> was implemented (AMTS  $< 7$  was an inclusion criterion for NCD), medical records were reviewed in all participants, and CT-scans were obtained where possible. Furthermore, in a number of subjects, the Mini-Mental State Exam (MMSE) Test<sup>24</sup> was implemented in addition to the AMTS. A score of  $< 24$  was an inclusion criterion for NCD.

The AMTS is currently one of the most accurate primary screening instruments to increase the probability of NCD<sup>25</sup>. 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<sup>22</sup>.

The control group was selected based on cognitive AMTS  $> 7$  and MMSE  $> 24$ , lack of major medical history, and normal CT-scan where possible. The cases and controls were matched based on age, gender, ethnicity, and residential district. The subjects' informed consent was obtained (from their guardians where necessary) and their identities remained confidential throughout the study. The research was approved by the Ethics Committee of the 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. All methods were performed in accordance with the relevant guidelines and regulations.

**Allele and genotype analysis of the *RASGEF1C* (GGC)-repeat.** Genomic DNA was obtained from peripheral blood using a standard salting out method. PCR reactions for the amplification of the *RASGEF1C* (GGC)-repeat were set up with the following primers.

**Forward:** GAGGGTGAAGCTGGGTTTGG.

**Reverse:** ACTCTAGCGGCTGAAAGAAG.

PCR reactions were carried out with a GC-TEMPase 2  $\times$  master mix (Amplicon) in a thermocycler (Peqlab-PEQStar) under the following conditions: touchdown PCR: 95 °C for 5 min, 20 cycles of denaturation at 95 °C for 45 s, annealing for 45 s at 67 °C ( $- 0.5$  decrease for each cycle) and extension at 72 °C for 1 min, and 30 cycles of denaturation at 95 °C for 40 s, annealing at 57 °C for 45 s and extension at 72 °C for 1 min, and a final extension at 72 °C for 10 min. Genotyping of every sample included in this study was performed following Sanger sequencing by the forward primer, using an ABI 3130 DNA sequencer (Supplementary Information 1 and 2).

**Analysis of the *RASGEF1C* (GGC)-repeat across vertebrates.** Ensembl 103 (<https://www.ensembl.org/index.html>) was used to analyze the interval between +1 and +100 of the TSS of the *RASGEF1C* in all the species in which this gene was annotated and the relevant region was sequenced. The CodonCode Aligner (<https://www.codoncode.com>) and Ensembl alignment programs (<http://www.ensembl.org>) were implemented for the sequence alignments across the species.

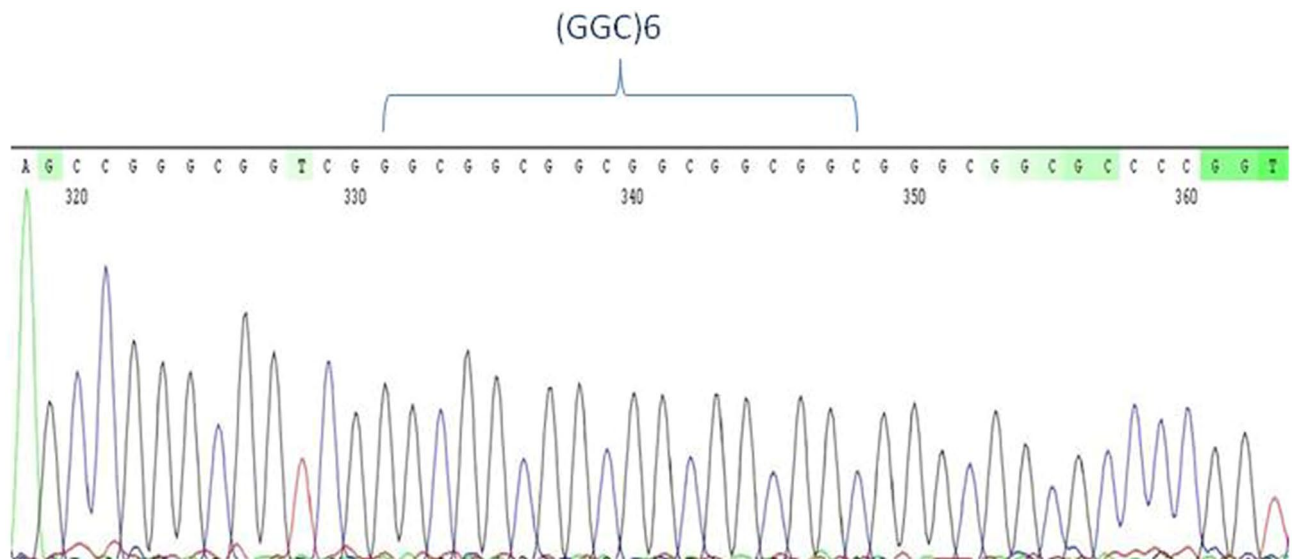
**Statistical analysis.** The P-values were calculated using the Two-by-Two Table of the OpenEpi calculator (<https://www.openepi.com/TwoByTwo/TwoByTwo.htm>)<sup>26</sup>.

**Statement of ethics.** The subjects' informed consent was obtained (from their guardians where necessary) and their identities remained confidential throughout the study. The research was approved by the Ethics Committee of the University 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.

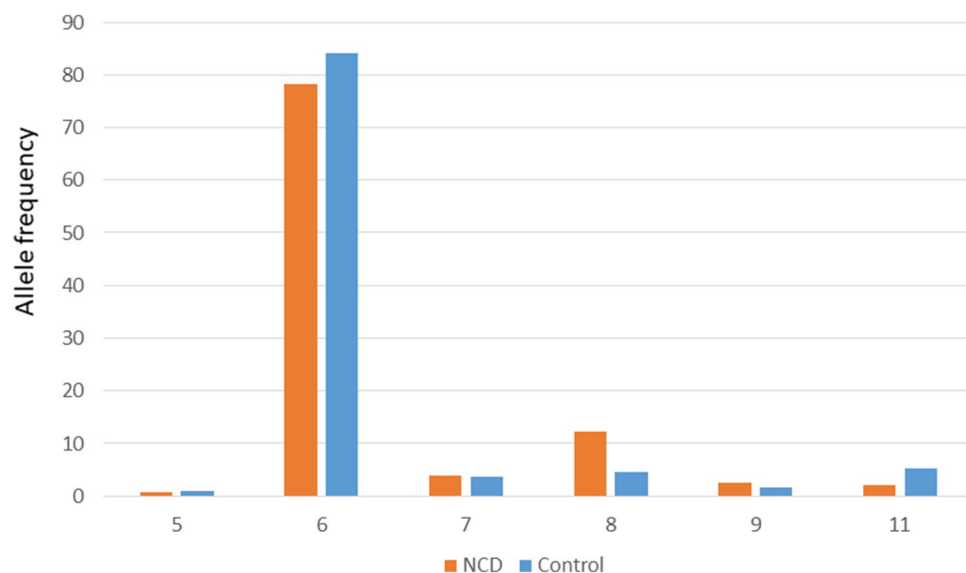
## Results

**Predominant abundance of the *RASGEF1C* (GGC)<sub>6</sub> in human..** We detected six alleles at 5, 6, 7, 8, 9, and 11-repeats, of which the predominant allele was the 6-repeat (Figs. 1 and 2). The frequency of (GGC)<sub>6</sub> was at 0.85 and 0.78 in the controls and NCD group, respectively (Fig. 2). At significantly lower frequencies, the 8 and 11 repeats ranked next in the NCD group and controls, respectively.

**Significant enrichment of divergent genotypes (genotypes that lacked the 6-repeat) in the NCD group.** We detected significant enrichment of genotypes that lacked the 6-repeat allele in the NCD group. Eleven out of 115 patients harbored such genotypes (Mid-P exact = 0.004) (Table 1, Figs. 3 and 4), whereas 3 out of 154 controls harbored those ( $p = 0.05$ ). The divergent genotypes consisted of the 7, 8, 9 and 11 repeat alleles, and heterozygous and homozygous genotypes were detected in that genotype compartment.



**Figure 1.** Electropherogram of the predominantly abundant allele at 6-repeats in the human *RASGEF1C* gene, in the context of a 6/6 genotype.



**Figure 2.** Allele frequency of the *RASGEF1C* (GGC)-repeat in NCD patients and controls. The 6-repeat was the predominant allele in both groups.

Among the divergent genotypes, in 5 patients (4% of the NCD group) we detected genotypes that were not detected in the control group (hence the term “disease-only”) (Mid-P exact = 0.007) (Table 1, Fig. 4).

Patients harboring the divergent genotypes spanned a wide age range, between 60 to 78 years, and revealed moderate to severe neurocognitive dysfunction. Possible diagnoses also varied, such as AD in patients 1, 5, 9, and 10 and vascular dementia in patients 2 and 11.

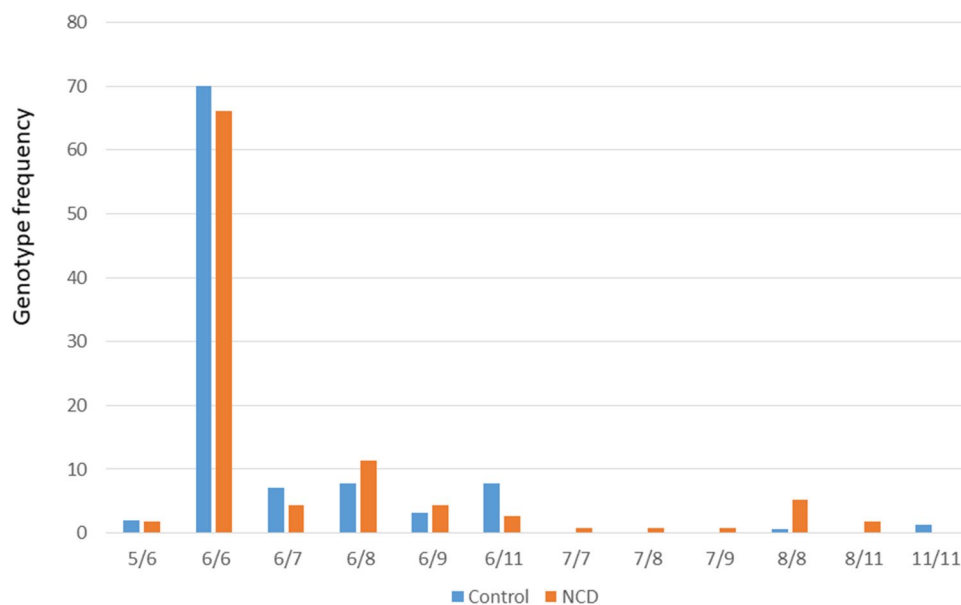
In line with a higher frequency of the 8-repeat in the NCD, we found a significant excess of the 8/8 genotype in this group in comparison to the control group (Mid-P exact = 0.01).

Although not statistically significant ( $p = 0.05$ ), two control individuals harbored the 11/11 genotype, which was not detected in the NCD group (Table 1). The frequency of the 11-repeat allele was also found to be higher in the controls vs. NCDs.

***RASGEF1C* (GGC)-repeat expanded specifically in primates, and was at maximum length in human.** Across all the species studied, the (GGC)-repeat was at maximum length in human. While in primates the minimum repeat length was 4-repeats (Fig. 5), the maximum length of (GGC)-repeat detectable in non-primates was 2-repeats (Fig. 6), indicating that this STR expanded specifically in primates.

Number	Sex	Age (years)	STR formula	AMTS	MMSE
<b>NCDs</b>					
1	F	87	7/7 <sup>a</sup>	2	–
2	F	62	7/8 <sup>a</sup>	5	19
3	M	72	7/9 <sup>a</sup>	2	16
4	F	70	8/8	4	–
5	F	77	8/8	4	–
6	F	67	8/8	4	–
7	F	73	8/8	5	–
8	F	63	8/8	5	–
9	M	79	8/8	1	–
10	F	79	8/11 <sup>a</sup>	3	18
11	M	60	8/11 <sup>a</sup>	5	21
<b>Controls</b>					
1	F	78	11/11	9	–
2	M	70	11/11	10	–
3	M	73	8/8	10	–

**Table 1.** NCD patients and controls harboring divergent genotypes (lacking the 6-repeat). *F* Female, *M* male. <sup>a</sup>Disease-only genotypes.

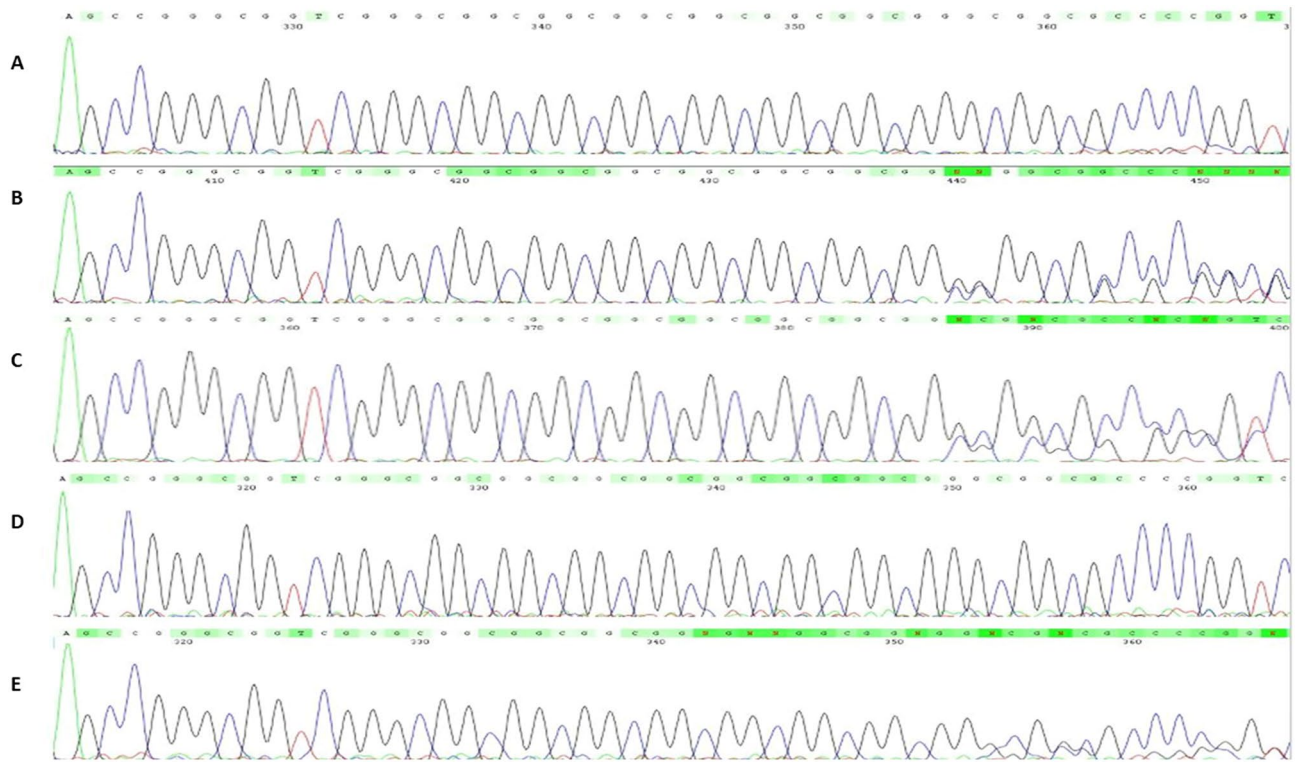


**Figure 3.** Genotype frequency of the *RASGEF1C* (GGC)-repeat in NCD patients and controls. While the 6/6 genotype was the predominant genotype in both groups, excess of divergent genotypes was detected in the NCD group.

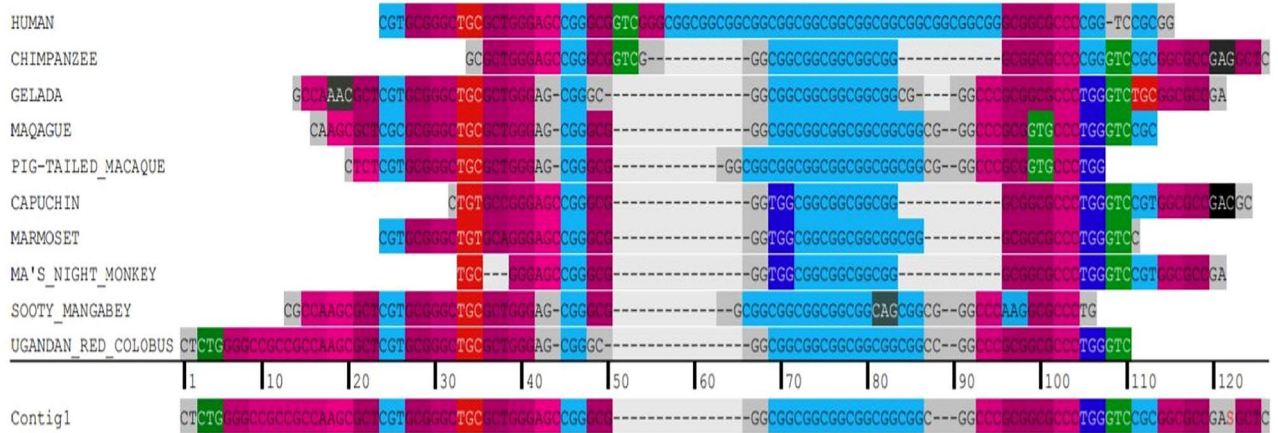
## Discussion

We propose that there is natural selection for the 6-repeat of the *RASGEF1C* (GGC)<sub>n</sub> in human. This proposition is not only based on the predominant abundance of the 6-repeat allele in the human subjects studied, but also the significant enrichment of divergent genotypes, lacking this allele in the NCD compartment. A number of divergent genotypes were detected in the NCD group only. Evidence of natural selection for an abundant allele in human has been previously reported by our group in the instance of the exceptionally long CA-repeat in the core promoter of the human *NHLH2* gene, and enrichment of genotypes lacking the predominantly abundant allele (the 21-repeat) in patients afflicted with late-onset NCD<sup>1</sup>. It is commonly assumed that genes influencing health in later life are not subject to natural selection. However, findings on the *APOE* alleles and several other NCD susceptibility loci<sup>27,28</sup> indicate that natural selection indeed happens on such alleles. *RASGEF1C* may also be linked to other yet to be identified phenotypes, which may impose natural selection at the (GGC)-repeat.

Based on the AceView database<sup>29</sup>, in comparison with several primates, the brain expression of *RASGEF1C* has the least quantile expression level in human (<https://www.ncbi.nlm.nih.gov/IEB/Research/AceView/>), which



**Figure 4.** Electropherogram of the non-6 genotypes (divergent genotypes) at the *RASGEF1C* (GGC) in NCD patients. (A) 7/7, (B) 7/8, (C) 7/9, (D) 8/8, (E) 8/11.



**Figure 5.** Sequence alignment of the *RASGEF1C* (GGC)-repeat across primate species, using CodonCode. The (GGC)-repeat was at maximum length in human.

coincides with maximum length of the (GGC)-repeat in human vs. all other primates. A (GGC)-repeat of similar length range in another gene, *Reelin* (*RELN*), which is critical to neurodevelopment in human, shows declining luciferase activity with increasing GGC repeat number<sup>30</sup>. However, it should be noted that the *RASGEF1C* (GGC)-repeat is among a number of other regulatory factors, which may also affect gene expression, and a link can only be established through future studies.

While aberrant repeat associated non-AUG translation (RAN), DNA hypermethylation, and polyglycine protein translation are established mechanisms linked to large (GGC)-repeat expansions (> 100 repeats) in a spectrum of neurological disorders in human<sup>5,31,32</sup>, the mechanisms underlying repeat selection at the range of the *RASGEF1C* (GGC)-repeat and also its link to late-onset NCD need to be clarified in the future studies.

Despite the high prevalence and debilitating characteristics of late-onset NCDs, genetic studies in this group of disorders have resulted in a number of genes with mild to modest effect for the most part<sup>33</sup>. In a novel approach, we selected the patients group based on late-onset NCD as an entity, without differentiating the NCD subtypes.





15. Annear, D. J. *et al.* Abundance of polymorphic CGG repeats in the human genome suggest a broad involvement in neurological disease. *Sci. Rep.* **11**(1), 1–11 (2021).
16. Jiao, B. *et al.* Identification of expanded repeats in NOTCH2NLC in neurodegenerative dementias. *Neurobiol. Aging*, **89**, 1421.e1–e7 (2020).
17. LaCroix, A. J. *et al.* GGC repeat expansion and exon 1 methylation of XYLT1 is a common pathogenic variant in Baratela-Scott syndrome. *Am. J. Hum. Genet.* **104**(1), 35–44 (2019).
18. Ma, D. *et al.* Association of NOTCH2NLC repeat expansions with Parkinson disease. *JAMA Neurol.* **77**(12), 1559–1563 (2020).
19. Sone, J. *et al.* Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. *Nat. Genet.* **51**(8), 1215–1221 (2019).
20. Ajjugal, Y., Kolimi, N. & Rathinavelan, T. Secondary structural choice of DNA and RNA associated with CGG/CCG trinucleotide repeat expansion rationalizes the RNA misprocessing in FXTAS. *Sci. Rep.* **11**(1), 1–17 (2021).
21. Kumutponpanich, T. *et al.* Clinicopathologic features of oculopharyngodistal myopathy with LRP12 CGG repeat expansions compared with other oculopharyngodistal myopathy subtypes. *JAMA Neurol.* **78**(7), 853–863. <https://doi.org/10.1001/jamaneurol.2021.1509> (2021).
22. Foroughan, M. *et al.* Validity and reliability of a abbreviated Mental TEst Score (AMTS) among older Iranian. *Psychogeriatrics* **17**(6), 460–465 (2017).
23. Hodkinson, H. Evaluation of a mental test score for assessment of mental impairment in the elderly. *Age Ageing* **1**(4), 233–238 (1972).
24. Folstein, M. A practical method for grading the cognitive state of patients for the children. *J. Psychiatr. Res.* **12**, 189–198 (1975).
25. Carpenter, C. R. *et al.* Accuracy of dementia screening instruments in emergency medicine: A diagnostic meta-analysis. *Acad. Emerg. Med. Off. J. Soc. Acad. Emerg. Med* **26**(2), 226–245 (2019).
26. Sullivan, K. M., Dean, A. & Soe, M. M. OpenEpi: A web-based epidemiologic and statistical calculator for public health. *Public Health Rep.* **124**(3), 471–474. <https://doi.org/10.1177/003335490912400320> (2009).
27. Drenos, F. & Kirkwood, T. B. Selection on alleles affecting human longevity and late-life disease: the example of apolipoprotein E. *PLoS One*. **5**(4), e10022 (2010).
28. Raj, T. *et al.* Alzheimer disease susceptibility loci: Evidence for a protein network under natural selection. *Am. J. Hum. Genet.* **90**(4), 720–726 (2012).
29. Thierry-Mieg, D. & Thierry-Mieg, J. AceView: A comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol.* **7**(1), 1–14 (2006).
30. Persico, A. M., Levitt, P. & Pimenta, A. F. Polymorphic GGC repeat differentially regulates human reelin gene expression levels. *J. Neural Transm. (Vienna)*. **113**(10), 1373–1382. <https://doi.org/10.1007/s00702-006-0441-6> (2006) (**Epub 2006 Apr 11**).
31. Sutcliffe, J. S. *et al.* DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum. Mol. Genet.* **1**(6), 397–400. <https://doi.org/10.1093/hmg/1.6.397> (1992).
32. Boivin, M. *et al.* Translation of GGC repeat expansions into a toxic polyglycine protein in NIID defines a novel class of human genetic disorders: The polyG diseases. *Neuron* **109**(11), 1825–1835.e5. <https://doi.org/10.1016/j.neuron.2021.03.038> (2021) (**Epub 2021 Apr 21**).
33. de Frutos-Lucas, J. *et al.* Does APOE genotype moderate the relationship between physical activity, brain health and dementia risk? A systematic review. *Ageing Res. Rev.* **64**, 101173 (2020).
34. Karantzoulis, S. & Galvin, J. E. Distinguishing Alzheimer's disease from other major forms of dementia. *Expert Rev. Neurother.* **11**(11), 1579–1591 (2011).
35. Lin, Y.-F. *et al.* Genetic overlap between vascular pathologies and Alzheimer's dementia and potential causal mechanisms. *Alzheimers Dement.* **15**(1), 65–75 (2019).
36. Noori, A., Mezlini, A. M., Hyman, B. T., Serrano-Pozo, A. & Das, S. Systematic review and meta-analysis of human Transcriptomics reveals Neuroinflammation, deficient energy metabolism, and Proteostasis failure across Neurodegeneration. *Neurobiol. Dis.* **167**, 105225 (2020).

## Acknowledgements

We wish to thank all the participants taking part in this research. This research was funded by the University of Social Welfare and Rehabilitation Sciences.

## Author contributions

Z.J. and S.K. performed the laboratory experiments. H.A. collected the clinical samples and their information. H.R. KH. and A.D. contributed to data collection and co-ordination. M.O. conceived and supervised the project, and wrote the manuscript. All authors reviewed the manuscript.

## Funding

This research was funded by the University of Social Welfare and Rehabilitation Sciences, Tehran, Iran.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-98725-y>.

**Correspondence** and requests for materials should be addressed to H.A. or M.O.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021