



OPEN

Novel implications of a strictly monomorphic (GCC) repeat in the human PRKACB gene

Safoura Khamse^{1,6}, Zahra Jafarian^{1,6}, Ali Bozorgmehr^{2✉}, Mostafa Tavakoli³, Hossein Afshar¹, Maryam Keshavarz⁴, Razieh Moayed⁵ & Mina Ohadi^{1✉}

PRKACB (Protein Kinase CAMP-Activated Catalytic Subunit Beta) is predominantly expressed in the brain, and regulation of this gene links to neuroprotective effects against tau and A β -induced toxicity. Here we studied a (GCC)-repeat spanning the core promoter and 5' UTR of this gene in 300 human subjects, consisting of late-onset neurocognitive disorder (NCD) (N = 150) and controls (N = 150). We also implemented several models to study the impact of this repeat on the three-dimensional (3D) structure of DNA. While the *PRKACB* (GCC)-repeat was strictly monomorphic at 7-repeats, we detected two 7/8 genotypes only in the NCD group. In all examined models, the (GCC)7 and its periodicals had the least range of divergence variation on the 3D structure of DNA in comparison to the 8-repeat periodicals and several hypothetical repeat lengths. A similar inert effect on the 3D structure was not detected in other classes of short tandem repeats (STRs) such as GA and CA repeats. In conclusion, we report monomorphism of a long (GCC)-repeat in the *PRKACB* gene in human, its inert effect on DNA structure, and enriched divergence in late-onset NCD. This is the first indication of natural selection for a monomorphic (GCC)-repeat, which probably evolved to function as an "epigenetic knob", without changing the regional DNA structure.

Abbreviations

| | |
|--------|--|
| 3D | Three-dimensional |
| NCD | Neurocognitive disorder |
| PRKACB | Protein Kinase CAMP-Activated Catalytic Subunit Beta |
| STR | Short tandem repeat |
| TSS | Transcription start site |
| UTR | Untranslated region |

Short tandem repeats (STRs) are the most polymorphic genetic elements in the vertebrate genomes. Because of their polymorphic nature and plasticity, these elements provide an efficient source of variation at the inter- and intraspecies levels^{1–6}. Accumulating evidence indicates that certain STRs may be associated with selective advantage in human and other species^{7–10}. Among the long STRs spanning the core promoter and 5' untranslated region (UTR) of human protein-coding genes, the protein kinase cAMP-activated catalytic subunit beta (*PRKACB*) gene contains a (GCC)-repeat of 7-repeats¹¹. Across human tissues, *PRKACB* has the highest level of expression in the brain¹² (<https://www.proteinatlas.org/ENSG00000142875-PRKACB/tissue>). Furthermore, in comparison with fourteen non-human primates, the brain expression of this gene reaches highest quantity in human (<https://www.ncbi.nlm.nih.gov/iebr/research/acemby/av.cgi?db=human&term=PRKACB&submit=Go>). The kinase encoded by *PRKACB* is involved in tau phosphorylation at Alzheimer's disease (AD)-related sites, and regulation of this gene has been linked to neuroprotective effects against tau and A β -induced toxicity^{13,14}.

The length of the *PRKACB* (GCC)-repeat in human is among the rare long lengths for this type of STR⁷. We hypothesized that this length may be of selective advantage to human. In view of the pathways in which *PRKACB*

¹Iranian Research Center on Aging, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. ²Research Center for Addiction and Risky Behaviors, Iran University of Medical Sciences, Tehran, Iran. ³Department of Applied Mathematics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, P.O. Box 1159, Mashhad 91775, Iran. ⁴Translational Biogerontology, German Center for Neurodegenerative Diseases (Deutsches Zentrum Für Neurodegenerative Erkrankungen, (DZNE)), Sigmund-Freud-str. 27, Bonn, Germany. ⁵Department of Economics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. ⁶These authors contributed equally: Safoura Khamse and Zahra Jafarian. ✉email: bozorgmehr.a@iums.ac.ir; mi.ohadi@uswr.ac.ir

is involved in, this (GCC)-repeat may have a role in the higher order brain functions, such as cognition. Here, we studied the (GCC)-repeat in a sample of human subjects, consisting of late-onset neurocognitive disorder (NCD) and controls, and implemented several models to analyze its impact on DNA three-dimensional (3D) structure.

Materials and methods

Subjects. Three hundred unrelated Iranian subjects, consisting of late-onset NCD patients (N=150) and controls (N=150) were recruited from the provinces of Tehran, Qazvin, and Rasht. In each NCD case, the Persian version of the Abbreviated Mental Test Score (AMTS)^{15,16} 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. The AMTS is currently one of the most accurate primary screening instruments to increase the probability of NCD¹⁷, and the Persian version of the AMTS is a valid cognitive assessment tool for older Iranian adults, which can be used for NCD screening in Iran¹⁵. The control group was selected based on cognitive AMTS of >7, lack of major medical history, and CT-scan where possible. The cases and controls were matched based on age, gender, 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 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. All methods were performed in accordance with the relevant guidelines and regulations.

Allele and genotype analysis of the *PRKACB* (GCC)-repeat. Genomic DNA was obtained from peripheral blood using a standard salting out method. PCR reactions for the amplification of the *PRKACB* (GCC)-repeat were set up with the following primers Forward: CGCCTGCGAAGATACAGTC, Reverse: CAACCTCACCGCTCTCCAC.

PCR reactions were carried out using High GC buffer (30% of the final volume per PCR reaction) in a thermocycler (PEQSTAR, model: PEQLAB) under the following conditions: 95 °C for 4 min, 35 cycles of denaturation at 94 °C for 30 s, annealing for 30 s at 58 °C and extension at 72 °C for 40 s, and a final extension of 72 °C for 5 min. All the samples included in this study were sequenced by the forward primer, using an ABI PRISM 3130 DNA sequencer. The divergent genotypes were sequenced by the reverse primer as well (Supplementary 1).

Analysis of the *PRKACB* (GCC)7 across vertebrates. The interval between +1 and +100 of the transcription start site (TSS) of the *PRKACB* was searched across several orders of vertebrates based on the Ensembl Release 102 (<https://asia.ensembl.org/index.html>). Alignment was performed using CodonCode Aligner 9.0.1.

In silico DNA reconstruction of various (GCC)-repeat lengths in human. The DNA 3D structure across the human *PRKACB* (GCC)-repeat was studied using several models, including AA-Wedge, DNase, Bolshoy, Cacchione, and Calladine (Supplementary 2)^{18–23}. For each reconstruction, the 100 nucleotide flanking sequences to the repeats were also included. As an instance, based on the twist, roll and tilt, the AA-Wedge model predicts experimental A-tract curvature as measured by gel retardation and cyclization kinetics¹⁸. In comparison to a number of existing DNA predicting models, such as the Crothers, Dickerson, Jernigan, Tung-Harvey, and Zhurkin, the AA-Wedge model has been reported as the most consistent and accurate¹⁹. Following obtaining the coordinates of the nucleotides in a 3D space, the DNA structure was visualized using plot3D package in R software²⁴.

Furthermore, three additional STRs, *SMAD9* (GCC)-STR, *RASGEF1* (GGC)-STR, and *GPM6B* (GA)-STR, were reconstructed according to the AA-Wedge model^{18,19}. The additional genes were studied as sample, for comparison between GCC and non-GCC STR effects.

Divergence calculation across DNA constructs of various periodical (GCC)-repeat lengths. Let (x_i, y_i, z_i) for $i = 1, 2, \dots, n$ be coordinates of points for each repeat, in the first step we scaled these coordinates as follows

$$\begin{aligned} X_i &= \frac{x_i - \min(x_i)}{(x_i) - \min(x_i)} & i = 1, 2, \dots, n \\ Y_i &= \frac{y_i - \min(y_i)}{(y_i) - \min(y_i)} & i = 1, 2, \dots, n \\ Z_i &= \frac{z_i - \min(z_i)}{(z_i) - \min(z_i)} & i = 1, 2, \dots, n \end{aligned}$$

Parametric equations of a segment line that passes through (X_i, Y_i, Z_i) 's were

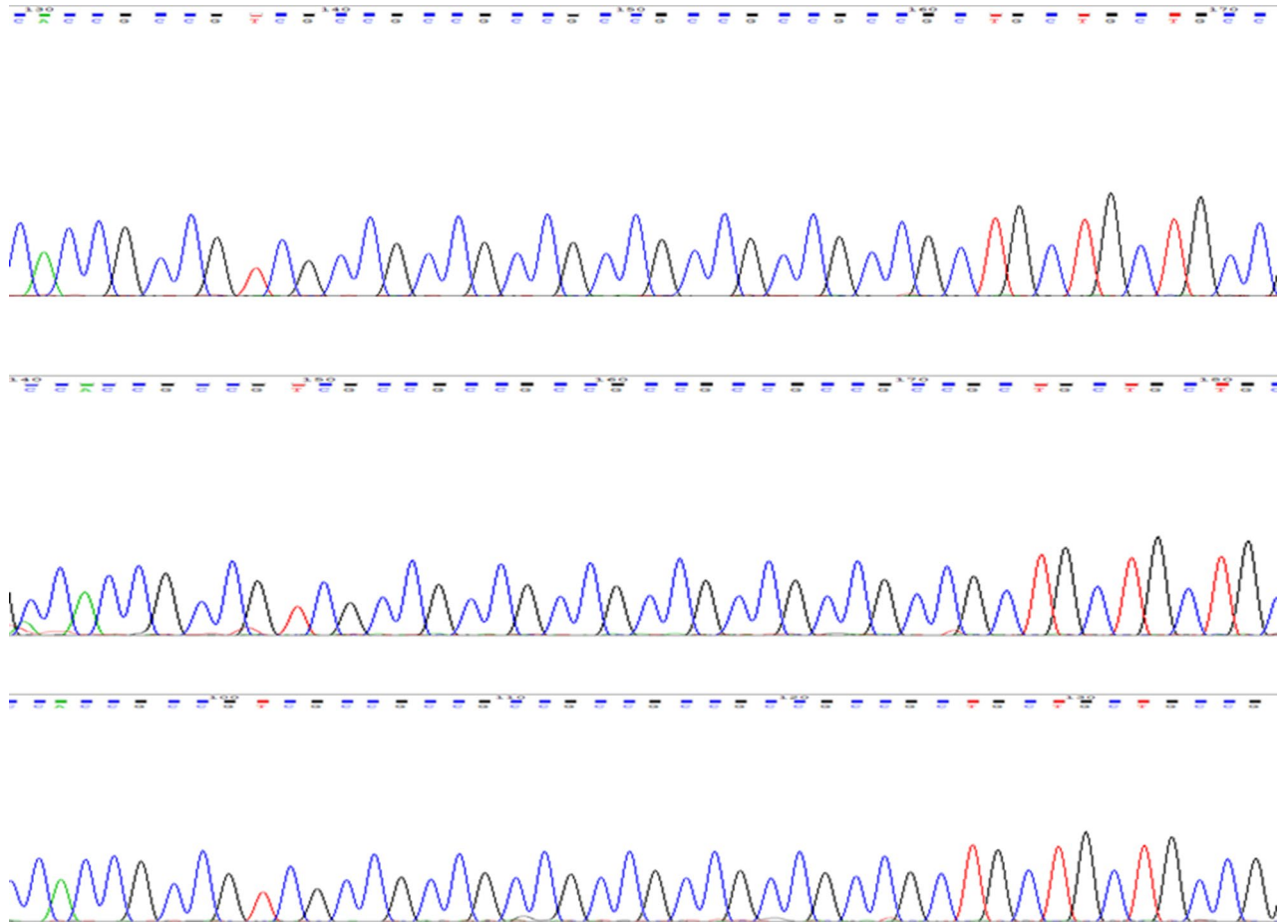


Figure 1. Strict monomorphism of the PRKACB (GCC)-repeat, at 7-repeats, in human. Only four samples are represented as examples.

$$\begin{aligned}
 \{X(t) &= \sum_{i=1}^{n-1} (X_i - (t(n-1) - (i-1))(X_{i+1} - X_i)I_{\left[\frac{i-1}{n-1}, \frac{i}{n-1}\right)}(t)Y(t) \\
 &= \sum_{i=1}^{n-1} (Y_i - (t(n-1) - (i-1))(Y_{i+1} - Y_i)I_{\left[\frac{i-1}{n-1}, \frac{i}{n-1}\right)}(t)Z(t) \\
 &= \sum_{i=1}^{n-1} (Z_i - (t(n-1) - (i-1))(Z_{i+1} - Z_i)I_{\left[\frac{i-1}{n-1}, \frac{i}{n-1}\right)}(t) \quad t \in [0, 1],
 \end{aligned}$$

where $I_{[a,b)}(t) = 1$ if $t \in [a, b)$ and $I_{[a,b)}(t) = 0$ if $t \notin [a, b)$. To measure similarity between two scaled lines, we defined integrated Euclidean distance (IED). Let $(X(t), Y(t), Z(t))$ and $(X'(t), Y'(t), Z'(t))$ for $t \in [0, 1]$ are two lines then IED is as follows

$$IED = \int_0^1 \sqrt{(X(t) - X'(t))^2 + (Y(t) - Y'(t))^2 + (Z(t) - Z'(t))^2} dt.$$

i,j)-th elements revealed the divergence scores obtained from the above method between the i-th and j-th diagrams.

The accuracy of data was validated by two-by-two comparison of constructs of identical lengths.

Gene network reconstruction. The STRING database (<https://string-db.org>) was used to find the interactions of PRKACB with other genes. STRING is a biological database of known molecular interactions containing information from experimental, computationally predicted, and public text collections²⁵. The minimum required interaction score was set at 0.7, and the maximum number of interactions was set at 20. To ensure the most important and reliable interactions, we selected only data from experimental studies as the active

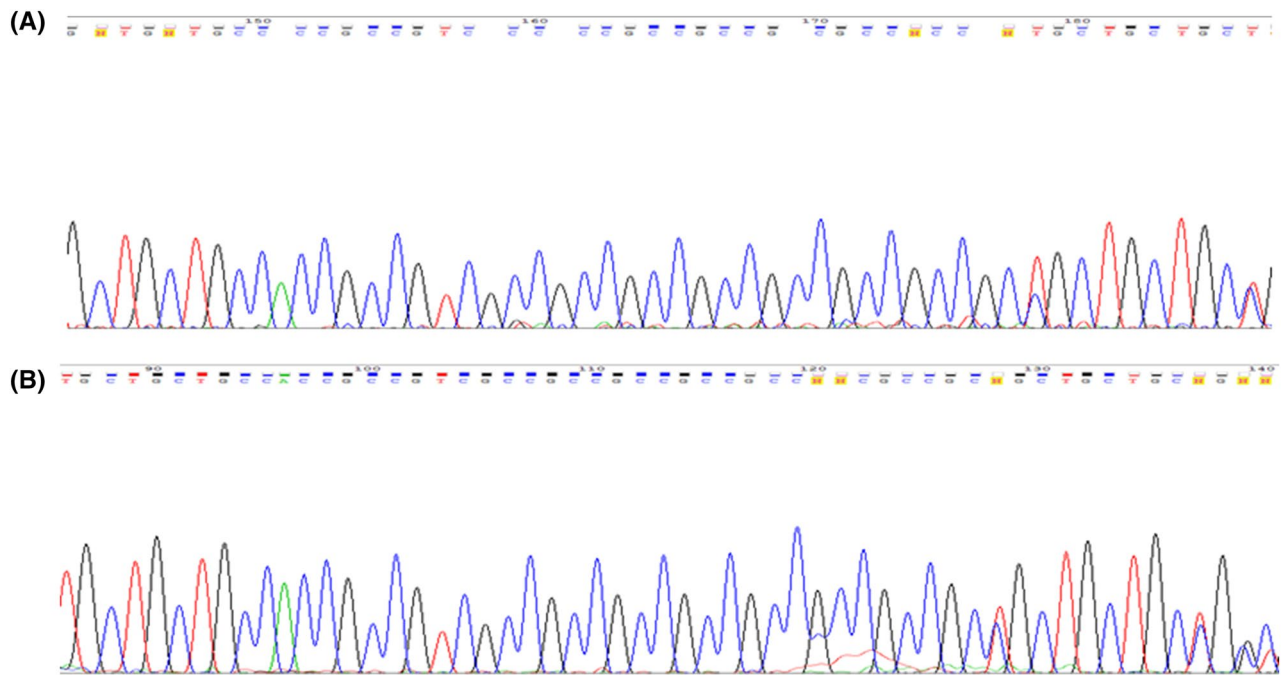


Figure 2. Two cases of late-onset NCD with 7/8 genotypes.

interaction sources. Subsequently, the interactive network was reconstructed, using Cytoscape version 3.8.2 and according to the interactions found²⁶.

Results

Predominant monomorphism of the *PRKACB* (GCC)-repeat in human, at 7-repeats, and divergence from monomorphism in two patients afflicted with late-onset NCD. We sequenced the *PRKACB* (GCC)-repeat in 300 human subjects, and found strict monomorphism of this STR at 7-repeats in this human sample (Fig. 1). Exceptions included two 7/8 genotypes in two patients with late-onset NCD (Fig. 2). The two patients harboring the 7/8 genotypes were females of 78 (Patient 1) and 83 (Patient 2) years of age and AMTS of 4 and 3, respectively.

In patient 1 the progression of neurocognitive symptoms was gradual, strengthening the possible diagnosis of AD. In patient 2, symptoms of neurocognitive impairment including aphonia and dementia occurred abruptly and subsequent to a cerebrovascular accident two years before the interview, strengthening the possible diagnosis of vascular dementia. Other causes of late-onset neurological disorders were ruled out in both patients by neurologists.

Status of the *PRKACB* (GCC)-repeat across vertebrates. The *PRKACB* (GCC)-repeat was highly conserved in numerous orders of vertebrates (Table 1). This repeat probably emerged in Birds and Reptiles, as we did not detect it in Amphibians, Fish, and other eukaryotes. In human, the interval encompassing the +1 to +100 to the TSS contains two (GCC)-repeats of (GCC)3 and (GCC)7 formula (Table 1). We detected a long ancestral trace of (GCC)-repeats, identifiable by triplet nucleotides of (GCC) or non-(GCC) in various species (Fig. 3). This long trace of (GCC) resulted in complex and unique (GCC) blocks in each species.

In silico reconstruction of the human *PRKACB* 5' UTR sequence encompassing the (GCC)-repeat. We detected only two alleles in the human population studied (7 and 8-repeats). We reconstructed DNA for those repeat lengths and their periodicals in order to analyze their effect on the 3D structure of the region. Across all studied models, the (GCC)7 periodicals e.g. (0, 7, 14), (1, 8, 15), (2, 9, 16), had a significantly less effect on the regional DNA structure in comparison to the 8-repeat periodicals e.g. (0, 8, 16), (1, 9, 17), (2, 10, 18) (Fig. 4, Table 2, Supplementary 2). While the 7-repeat periodicals were strictly consistent in respect of having the least amount of divergence variation, all other repeat length periodicals, including the 8-repeat, were noncontinuous and interrupted by high divergence scores.

To examine the specificity of the *PRKACB* (GCC)-repeat inert effect, we performed DNA reconstruction of additional GCC/GGC-repeats in the 5' UTR of the *SMAD9* and *RASGEF1C* genes, and also a non-GCC STR, such as a GA-repeat in the 5' UTR of human *GPM6B* (Supplementary 3). Strikingly, while the reconstructed 3D structures were significantly divergent at various lengths in the instance of the *GPM6B* 5' UTR GA-repeat, the *SMAD9* and *RASGEF1C* 3D structures were almost identical for all repeat lengths studied. We also studied the interactive effect of divergent non-repeat flanking sequences with the (GCC)-repeats by studying a different species than human, such as capuchin. The inert effect of the (GCC)-block was consistent, regardless of the differential flanking non-repeat sequences in capuchin versus human.

| Species | Transcript ID | STR formula | Species | Transcript ID | STR formula |
|--------------------------|----------------------|----------------|--------------------------|------------------------|---------------|
| Primates | | | Leopard | ENSPPT00000024708.1 | (GCC)3 |
| Human | ENST00000370689.6 | (GCC)3 (GCC)7 | Dingo | – | – |
| Bonobo | – | – | Shrew | – | – |
| Chimpanzee | ENSPTRT00000109165.1 | (GCC)3 (GCC)4 | | | |
| Gorilla | ENSGGOT00000045976.1 | (GCC)3 (GCC)3 | Lamprey | – | – |
| Orangutan | – | – | Coelacanth | – | – |
| Gibbon | ENSNLET00000012651.2 | (GCC)2 | Elephant shark | ENSCMIT00000019194.1 | (GCC)2 |
| Crab-eating macaque | ENSMFAT00000033375.1 | (GCC)2 (GCC)13 | Hagfish | – | – |
| Macaque | ENSMUT00000076404.2 | (GCC)2 (GCC)10 | Afrotheria | | |
| Golden snub-nosed monkey | ENSRROT00000025377.1 | (GCC)2 (GCC)8 | Elephant | – | – |
| Gelada | ENSTGET00000029662.1 | (GCC)2 (GCC)8 | Hyrax | – | – |
| Mouse lemur | ENSMICT00000059697.1 | (GCC)5 | Lesser hedgehog tenrec | – | – |
| Vervet-AGM | – | – | Xenarthra | | |
| Marmoset | ENSCJAT00000115300.1 | (GCC)3 (GCC)4 | Armadillo | – | – |
| Sooty mangabey | ENSCATT00000063374.1 | (GCC)2 (GCC)3 | Sloth | – | – |
| Drill | ENSMLET00000056274.1 | (GCC)7 (GCC)1 | Other mammals | | |
| Pig-tailed macaque | ENSMNET00000056095.1 | (GCC)2 (GCC)6 | Common wombat | – | – |
| Angola colobus | – | – | Koala | – | – |
| Ugandan red Colobus | ENSPTET00000046092.1 | (GCC)11 | Opossum | ENSMODT00000083179.1 | (GCC)2 |
| Black snub-nosed monkey | ENSRBIT00000058758.1 | (GCC)2 (GCC)6 | Platypus | ENSOANT00000057701.1 | (GCC)2 |
| Ma's night monkey | ENSANAT00000031038.1 | (GCC)3 (GCC)2 | Birds and reptiles | | |
| Capuchin | ENSCCAT00000025557.1 | (GCC)3 (GCC)8 | Bengalese finch | ENSLSDT00000018202.1 | (GCC)2 (GCC)3 |
| Bolivian squirrel monkey | ENSSBOT00000050020.1 | (GCC)2 | Blue tit | – | – |
| Tarsier | – | – | Burrowing owl | – | – |
| Greater bamboo lemur | ENSPSMT00000037282.1 | (GCC)5 | Chicken | ENSGALT00000062352.2 | (GCC)3 |
| Coquerel's sifaka | ENSPCOT00000017095.1 | (GCC)7 | Common canary | – | – |
| Bushbaby | – | – | Golden eagle | ENSACCT00020005753.1 | (GCC)4 |
| Rodent | | | Kakapo | ENSSHBT0000005016149.1 | (GCC)4 |
| Mouse | ENSMUST00000102515.9 | (GCC)3 (GCC)4 | Anole lizard | ENSACAT00000030613.1 | (GCC)4 |
| Long-tailed chinchilla | ENSCLAT00000008409.1 | (GCC)6 | Eastern brown snake | ENSPTXT00000009104.1 | (GCC)2 |
| Kangaroo rat | ENSDORT00000033480.1 | (GCC)3 (GCC)7 | Amphibians | | |
| Naked mole-rat female | ENSHGLT00000016208.1 | (GCC)5 | Tropical clawed frog | – | – |
| Rabbit | ENSOCUT00000051257.1 | (GCC)5 (GCC)10 | Fish | | |
| Laurasiatheria | | | Asian bonytongue | – | – |
| Cat | ENSFCAT00000050575.2 | (GCC)3 | Zebrafish | – | – |
| Cow | ENSBTAT00000074614.1 | (GCC)4 (GCC)5 | Common crap | – | – |
| Goat | – | – | Zig-zag eel | – | – |
| Blue whale | ENSBMST00010016111.1 | (GCC)3 | Other eukaryotes | | |
| Horse | ENSECAT00000063662.1 | (GCC)8 | Saccharomyces cerevisiae | – | – |
| Siberian musk deer | ENSMST00000013945.1 | (GCC)6 | Drosophila melanogaster | – | – |
| Red Fox | ENSVVUT00000055345.1 | (GCC)2 | Caenorhabditis elegans | – | – |
| PIG | | | Other cordates | | |
| Dog | – | – | C.intestinalis | – | – |
| Panda | – | – | C.savignyi | – | – |

Table 1. Status of the *PRKACB* (GCC)-repeat across vertebrates.

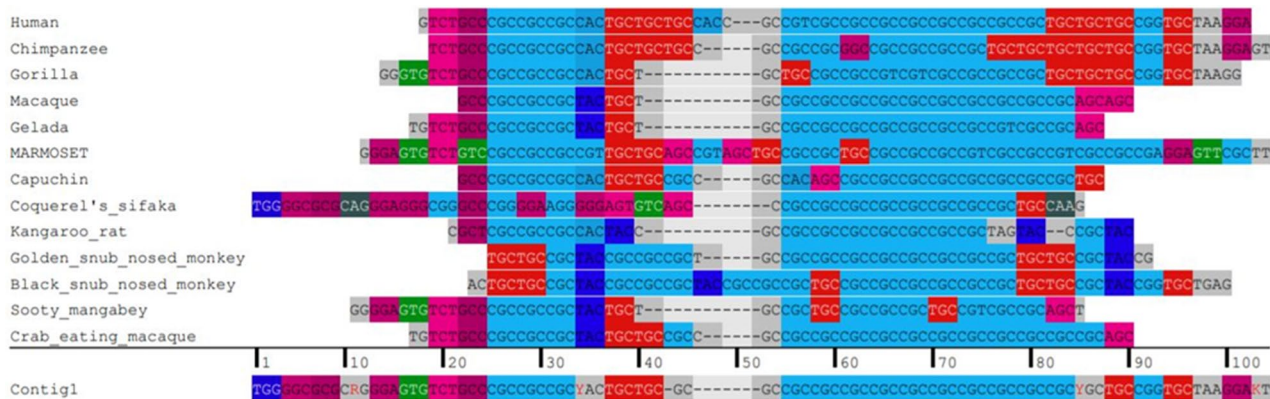


Figure 3. Sequence alignment of the *PRKACB* region encompassing the (GCC)-repeat in various primates. Each species revealed complex and unique stretches of (GCC)-repeats. Contig1 indicates the possible ancestral sequence across the selected species. Triplet nucleotides which may be remnants of GCC sequences are detectable in the ancestral sequence.

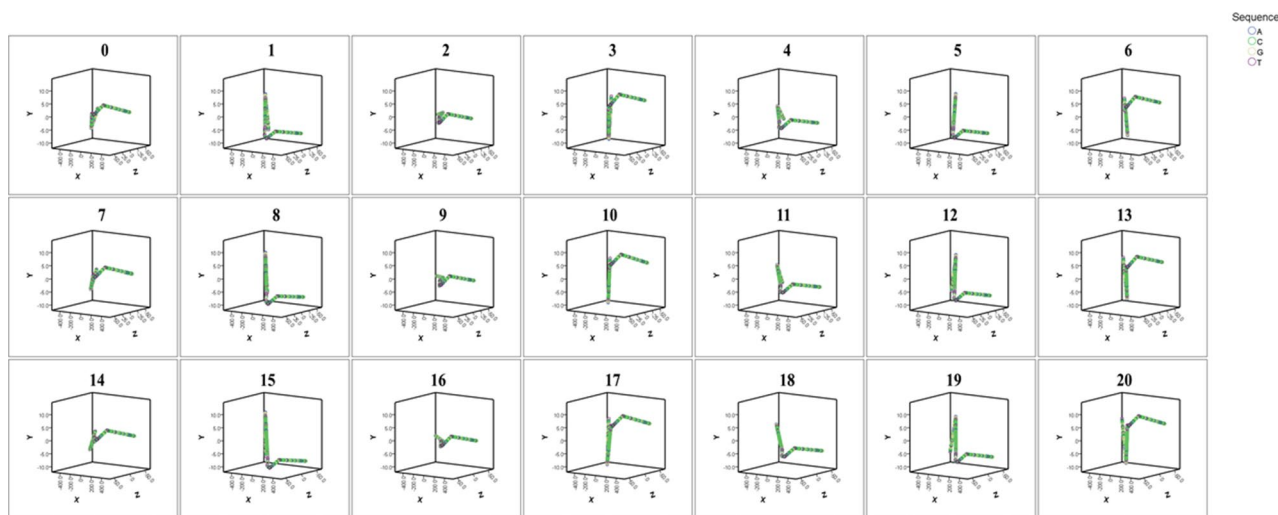


Figure 4. 3D construction of human DNA encompassing the *PRKACB* (GCC)-repeat based on the AA-Wedge model. Striking similarity was detected in repeats where subtraction or addition of 7-repeats was included.

Gene network reconstruction. Based on the available experimental evidence, the reconstructed interactive network consisted of 16 nodes and 45 edges (Fig. 5). Within this network, *PRKACB* interacts with genes of other subunits of the cAMP-dependent protein kinase complex, as well as other genes such as *MAPT*, *FOXO1*, and *PDE5A*, which are central to maintaining cell structure and function.

Discussion

Because of containing internal CpG dinucleotides, susceptible to methylation and subsequent base substitutions, (GCC)-repeats are potential mutation hotspots^{27–29}, and this may be the reason why this class of repeats do not expand beyond certain lengths unless where selected, such as in the gene regulatory regions³⁰. Large expansions of (GCC)-repeats in the 5' UTR of a number of genes are associated with hypermethylation and intellectual disability³¹. Based on the Ensembl database, the *PRKACB* gene is in the 10th percentile of genes in respect of containing long (GCC)-repeats¹⁴. Here we show that this (GCC)-repeat is predominantly monomorphic in human at 7-repeats, and there may be enrichment of divergence from this monomorphism in late-onset NCD. We propose that there was selective advantage in human at this particular STR to reach to, and stabilize at that length. The above proposition is supported by the observations that mutations that have negative fitness consequences tend to be eliminated from the population³². This STR was found to be highly conserved across various mammalian species, indicating its important role in growth and development. In general, STRs near the TSSs of genes are often highly conserved, and distance from a STR to the nearest TSS is a good predictor of the STR conservation score³³.

We also found that (GCC)7 periodicals had the least effect on the 3D structure in contrast to the more significant effects that various other repeats inserted, which probably confers an inert effect to (GCC)7 without changing the structure of DNA or RNA. The above was deduced from the exceptionally continuous trend of

| Repeats | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 0 | 0.00 | 0.51 | 0.37 | 0.12 | 0.38 | 0.57 | 0.25 | 0.05 | 0.52 | 0.37 | 0.13 | 0.43 | 0.60 | 0.26 | 0.09 | 0.53 | 0.39 | 0.14 | 0.47 | 0.60 | 0.27 |
| 1 | 0.51 | 0.00 | 0.39 | 0.53 | 0.19 | 0.16 | 0.57 | 0.49 | 0.05 | 0.44 | 0.52 | 0.15 | 0.20 | 0.57 | 0.48 | 0.09 | 0.49 | 0.52 | 0.14 | 0.23 | 0.57 |
| 2 | 0.37 | 0.39 | 0.00 | 0.35 | 0.31 | 0.34 | 0.27 | 0.35 | 0.39 | 0.10 | 0.34 | 0.36 | 0.34 | 0.27 | 0.34 | 0.39 | 0.20 | 0.34 | 0.39 | 0.35 | 0.27 |
| 3 | 0.12 | 0.53 | 0.35 | 0.00 | 0.42 | 0.56 | 0.16 | 0.14 | 0.53 | 0.35 | 0.04 | 0.47 | 0.58 | 0.17 | 0.17 | 0.54 | 0.38 | 0.07 | 0.50 | 0.58 | 0.19 |
| 4 | 0.38 | 0.19 | 0.31 | 0.42 | 0.00 | 0.33 | 0.48 | 0.36 | 0.21 | 0.37 | 0.41 | 0.09 | 0.36 | 0.48 | 0.36 | 0.24 | 0.44 | 0.41 | 0.14 | 0.37 | 0.49 |
| 5 | 0.57 | 0.16 | 0.34 | 0.56 | 0.33 | 0.00 | 0.52 | 0.56 | 0.13 | 0.39 | 0.55 | 0.28 | 0.06 | 0.52 | 0.55 | 0.14 | 0.46 | 0.55 | 0.25 | 0.09 | 0.52 |
| 6 | 0.25 | 0.57 | 0.27 | 0.16 | 0.48 | 0.52 | 0.00 | 0.25 | 0.57 | 0.25 | 0.15 | 0.53 | 0.52 | 0.04 | 0.27 | 0.56 | 0.28 | 0.17 | 0.56 | 0.52 | 0.06 |
| 7 | 0.05 | 0.49 | 0.35 | 0.14 | 0.36 | 0.56 | 0.25 | 0.00 | 0.51 | 0.35 | 0.13 | 0.42 | 0.58 | 0.26 | 0.04 | 0.52 | 0.37 | 0.13 | 0.45 | 0.59 | 0.27 |
| 8 | 0.52 | 0.05 | 0.39 | 0.53 | 0.21 | 0.13 | 0.57 | 0.51 | 0.00 | 0.43 | 0.53 | 0.16 | 0.17 | 0.57 | 0.49 | 0.05 | 0.49 | 0.52 | 0.13 | 0.19 | 0.56 |
| 9 | 0.37 | 0.44 | 0.10 | 0.35 | 0.37 | 0.39 | 0.25 | 0.35 | 0.43 | 0.00 | 0.34 | 0.39 | 0.39 | 0.25 | 0.32 | 0.43 | 0.10 | 0.33 | 0.42 | 0.40 | 0.25 |
| 10 | 0.13 | 0.52 | 0.34 | 0.04 | 0.41 | 0.55 | 0.15 | 0.13 | 0.53 | 0.34 | 0.00 | 0.46 | 0.57 | 0.15 | 0.15 | 0.53 | 0.36 | 0.03 | 0.49 | 0.57 | 0.17 |
| 11 | 0.43 | 0.15 | 0.36 | 0.47 | 0.09 | 0.28 | 0.53 | 0.42 | 0.16 | 0.39 | 0.46 | 0.00 | 0.30 | 0.53 | 0.40 | 0.18 | 0.44 | 0.46 | 0.06 | 0.31 | 0.53 |
| 12 | 0.60 | 0.20 | 0.34 | 0.58 | 0.36 | 0.06 | 0.52 | 0.58 | 0.17 | 0.39 | 0.57 | 0.30 | 0.00 | 0.52 | 0.57 | 0.14 | 0.45 | 0.56 | 0.27 | 0.04 | 0.51 |
| 13 | 0.26 | 0.57 | 0.27 | 0.17 | 0.48 | 0.52 | 0.04 | 0.26 | 0.57 | 0.25 | 0.15 | 0.53 | 0.52 | 0.00 | 0.27 | 0.56 | 0.26 | 0.15 | 0.55 | 0.52 | 0.03 |
| 14 | 0.09 | 0.48 | 0.34 | 0.17 | 0.36 | 0.55 | 0.27 | 0.04 | 0.49 | 0.32 | 0.15 | 0.40 | 0.57 | 0.27 | 0.00 | 0.51 | 0.34 | 0.14 | 0.44 | 0.57 | 0.27 |
| 15 | 0.53 | 0.09 | 0.39 | 0.54 | 0.24 | 0.14 | 0.56 | 0.52 | 0.05 | 0.43 | 0.53 | 0.18 | 0.14 | 0.56 | 0.51 | 0.00 | 0.49 | 0.52 | 0.14 | 0.16 | 0.56 |
| 16 | 0.39 | 0.49 | 0.20 | 0.38 | 0.44 | 0.46 | 0.28 | 0.37 | 0.49 | 0.10 | 0.36 | 0.44 | 0.45 | 0.26 | 0.34 | 0.49 | 0.00 | 0.35 | 0.46 | 0.45 | 0.26 |
| 17 | 0.14 | 0.52 | 0.34 | 0.07 | 0.41 | 0.55 | 0.17 | 0.13 | 0.52 | 0.33 | 0.03 | 0.46 | 0.56 | 0.15 | 0.14 | 0.52 | 0.35 | 0.00 | 0.48 | 0.57 | 0.15 |
| 18 | 0.47 | 0.14 | 0.39 | 0.50 | 0.14 | 0.25 | 0.56 | 0.45 | 0.13 | 0.42 | 0.49 | 0.06 | 0.27 | 0.55 | 0.44 | 0.14 | 0.46 | 0.48 | 0.00 | 0.27 | 0.55 |
| 19 | 0.60 | 0.23 | 0.35 | 0.58 | 0.37 | 0.09 | 0.52 | 0.59 | 0.19 | 0.40 | 0.57 | 0.31 | 0.04 | 0.52 | 0.57 | 0.16 | 0.45 | 0.57 | 0.27 | 0.00 | 0.51 |
| 20 | 0.27 | 0.57 | 0.27 | 0.19 | 0.49 | 0.52 | 0.06 | 0.27 | 0.56 | 0.25 | 0.17 | 0.53 | 0.51 | 0.03 | 0.27 | 0.56 | 0.26 | 0.15 | 0.55 | 0.51 | 0.00 |

Table 2. Divergence scores of existing (7 and 8-repeat) and hypothetical *PRKACB* repeat periodicals in human. The highlighted areas represent the 30 percentile of the entire scores (the lowest 30% of the entire scores). The 7-repeat periodical (solid yellow cells) revealed continuous and strictly narrow divergence score variation in comparison to the 8-repeat and all other periodicals. Only the 7 and 8 repeats were detected in the human population studied.

those periodicals (uninterrupted by high divergence scores) in contrast to all other repeats. Interestingly, the divergence score between the 7 and 8 repeat (the latter was an allele of the two divergent genotypes) was among the highest in all examined models. While we also detected an inert effect of GCC/GGC repeats in two additional genes, *SMAD9* and *RASGEF1C*, this effect was not observed in the GA class of STRs, where various lengths of GA-repeats resulted in significant variation in the 3D structure of DNA in the present study and in a previously reported research in the instance of human *ZMYM3*¹¹. Significant divergence has also been observed with various CA-repeat lengths in the human *NHLH2* gene⁹. The above findings indicate that long GCC/GGC repeats might have evolved to exert epigenetic effects, as epigenetic knobs, without changing the structure of the regional DNA.

GCC motifs of STRs significantly overlap with G-quadruplex (G4) non-B structures³⁴. Recent research indicates that organisms may have evolutionarily developed G4 into a novel reversible and elaborate transcriptional regulatory mechanism benefiting multiple physiological activities of higher organisms^{35,36}.

PRKACB is involved in a number of physiological and pathological pathways, the outcomes of which directly affect cognition in human^{12,13,37–39}. Future studies are warranted to sequence the *PRKACB* (GCC)-repeat in a large number of patients with neurological disorders. Considering the pivotal role of *PRKACB* in the brain, CRISPR/Cas9 methods may also be ideal to edit this STR at the genomic level, and investigate differentiation of human stem cells into neural cells.

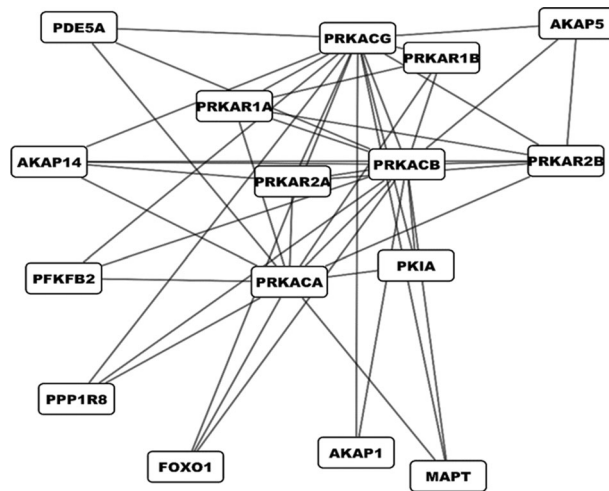


Figure 5. The interactive network reconstructed based on experimental evidence. The network consists of 16 nodes, including genes of the cAMP-dependent protein kinase subunits, as well as other structurally and functionally important genes such as MAPT, FOXO1, and PDE5A. The 16 genes interact with each other through 45 edges.

Conclusion

In conclusion, we report the first instance of predominant monomorphism of a long (GCC)-repeat in human and enriched divergent genotypes from this monomorphism in human disease. We also propose that (GCC)-repeats might have evolved in the genome as regulatory elements (such as epigenetic regulation) without dramatically changing the 3D structure of the regional DNA. Further experimental research is warranted to examine the proposed model.

Received: 16 April 2021; Accepted: 5 October 2021

Published online: 19 October 2021

References

- Jakubosky, D. *et al.* Properties of structural variants and short tandem repeats associated with gene expression and complex traits. *Nat. Commun.* **11**, 1–15 (2020).
- Fotsing, S. F. *et al.* The impact of short tandem repeat variation on gene expression. *Nat. Genet.* **51**, 1652–1659 (2019).
- Press, M. O., Hall, A. N., Morton, E. A. & Queitsch, C. Substitutions are boring: Some arguments about parallel mutations and high mutation rates. *Trends Genet.* **35**, 253–264 (2019).
- Valipour, E. *et al.* Polymorphic core promoter GA-repeats alter gene expression of the early embryonic developmental genes. *Gene* **531**, 175–179 (2013).
- Ohadi, M. *et al.* Core promoter short tandem repeats as evolutionary switch codes for primate speciation. *Am. J. Primatol.* **77**, 34–43 (2015).
- Mohammadparast, S., Bayat, H., Biglarian, A. & Ohadi, M. Exceptional expansion and conservation of a CT-repeat complex in the core promoter of PAXBP1 in primates. *Am. J. Primatol.* **76**, 747–756 (2014).
- Namdar-Aligoodarzi, P. *et al.* Exceptionally long 5' UTR short tandem repeats specifically linked to primates. *Gene* **569**(1), 88–94. <https://doi.org/10.1016/j.gene.2015.05.053> (2015) (Epub 2015 May 27 PMID: 26022613).
- Watts, P. C. *et al.* Stabilizing selection on microsatellite allele length at arginine vasopressin 1a receptor and oxytocin receptor loci. *Proc. R. Soc. B Biol. Sci.* **284**, 20171896 (2017).
- Afshar, H. *et al.* Evolving evidence on a link between the ZMYM3 exceptionally long GA-STR and human cognition. *Sci. Rep.* **10**, 1–7 (2020).
- Afshar, H. *et al.* Natural selection at the NHLH2 core promoter exceptionally long CA-repeat in human and disease-only genotypes in late-onset neurocognitive disorder. *Gerontology* **66**, 514–522 (2020).
- Alizadeh, F. *et al.* Disease-only alleles at the extreme ends of the human ZMYM3 exceptionally long 5' UTR short tandem repeat in bipolar disorder: A pilot study. *J. Affect. Disord.* **251**, 86–90 (2019).
- London, E. *et al.* The catalytic subunit β of PKA affects energy balance and catecholaminergic activity. *J. Endocr. Soc.* **3**, 1062–1078 (2019).
- Wang, L. *et al.* MicroRNA-200a-3p mediates neuroprotection in Alzheimer-related deficits and attenuates amyloid-beta overproduction and tau hyperphosphorylation via coregulating BACE1 and PRKACB. *Front. Pharmacol.* **10**, 806 (2019).
- Vázquez-Higuera, J. L. *et al.* Genetic variation in the tau kinases pathway may modify the risk and age at onset of Alzheimer's disease. *J. Alzheimers Dis.* **27**, 291–297 (2011).
- Foroughan, M. *et al.* Validity and reliability of Abbreviated Mental Test Score (AMTS) among older Iranian. *Psychogeriatr. Off. J. Jpn. Psychogeriatr. Soc.* **17**, 460–465 (2017).
- Hodkinson, H. Evaluation of a mental test score for assessment of mental impairment in the elderly. *Age Ageing* **1**, 233–238 (1972).
- Carpenter, C. R. *et al.* Accuracy of dementia screening instruments in emergency medicine: A diagnostic meta-analysis. *Acad. Emerg. Med.* **26**, 226–245 (2019).
- Tan, R. K. Z. & Harvey, S. C. A comparison of six DNA bending models. *J. Biomol. Struct. Dyn.* **5**, 497–512 (1987).
- Goodsell, D. S. & Dickerson, R. E. Bending and curvature calculations in B-DNA. *Nucleic Acids Res.* **22**, 5497–5503 (1994).

20. Lazarovici, A. *et al.* Probing DNA shape and methylation state on a genomic scale with DNase I. *Proc. Natl. Acad. Sci. U. S. A.* **110**(16), 6376–6381. <https://doi.org/10.1073/pnas.1216822110> (2013) (Epub 2013 Apr 1. PMID: 23576721; PMCID: PMC3631675).
21. Bolshoy, A., McNamara, P., Harrington, R. & Trifonov, E. Curved DNA without A-A: Experimental estimation of all 16 DNA wedge angles. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2312–2316 (1991).
22. Cacchione, S., De Santis, P., Foti, D., Palleschi, A. & Savino, M. Periodical polydeoxynucleotides and DNA curvature. *Biochemistry* **28**(22), 8706–8713 (1989).
23. Calladine, C. R., Drew, H. R. & McCall, M. J. The intrinsic curvature of DNA in solution. *J. Mol. Biol.* **201**(1), 127–137 (1988).
24. Soetaert, K. plot3Drgl: Tools for plotting 3-D and 2-D data in openGL.
25. Szklarczyk, D. *et al.* The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).
26. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
27. Jang, H. S., Shin, W. J., Lee, J. E. & Do, J. T. CpG and Non-CpG methylation in epigenetic gene regulation and brain function. *Genes (Basel)* **8**(6), 148. <https://doi.org/10.3390/genes8060148>. PMID:28545252;PMCID:PMC5485512 (2017).
28. Li, Q. S., Sun, Y. & Wang, T. Epigenome-wide association study of Alzheimer's disease replicates 22 differentially methylated positions and 30 differentially methylated regions. *Clin. Epigenet.* **12**, 1–14 (2020).
29. Fryxell, K. J. & Moon, W.-J. CpG mutation rates in the human genome are highly dependent on local GC content. *Mol. Biol. Evol.* **22**(3), 650–658. <https://doi.org/10.1093/molbev/msi043> (2005).
30. Sawaya, S. *et al.* Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. *PLoS ONE* **8**, e54710 (2013).
31. Lesca, G. *et al.* Clinical, cytogenetic, and molecular description of a FRAXE French family. *Psychiatr. Genet.* **13**, 43–46 (2003).
32. Gymrek, M., Willems, T., Reich, D. & Erlich, Y. Interpreting short tandem repeat variations in humans using mutational constraint. *Nat. Genet.* **49**, 1495 (2017).
33. Sawaya, S. M., Lennon, D., Buschiazzi, E., Gemmell, N. & Minin, V. N. Measuring microsatellite conservation in mammalian evolution with a phylogenetic birth–death model. *Genome Biol. Evol.* **4**, 636–647 (2012).
34. Spiegel, J., Adhikari, S. & Balasubramanian, S. The structure and function of DNA G-quadruplexes. *Trends Chem.* **2**, 123–136 (2020).
35. Niu, K. *et al.* Identification of LARK as a novel and conserved G-quadruplex binding protein in invertebrates and vertebrates. *Nucleic Acids Res.* **47**, 7306–7320 (2019).
36. Wu, F. *et al.* Genome-wide analysis of DNA G-quadruplex motifs across 37 species provides insights into G4 evolution. *Commun. Biol.* **4**, 1–11 (2021).
37. Wu, Y. Q. *et al.* Microarray analysis identifies key differentially expressed circular RNAs in aged mice with postoperative cognitive dysfunction. *Front Aging Neurosci.* **16**(13), 716383. <https://doi.org/10.3389/fnagi.2021.716383>. PMID:34483886;PMCID:PMC8415796 (2021).
38. Zhang, Z., Shen, Q., Wu, X., Zhang, D. & Xing, D. Activation of PKA/SIRT1 signaling pathway by photobiomodulation therapy reduces A β levels in Alzheimer's disease models. *Aging Cell* **19**, e13054. <https://doi.org/10.1111/acer.13054> (2020).
39. Dagda, R. K. & Das, B. T. Role of protein kinase a in regulating mitochondrial function and neuronal development: Implications to neurodegenerative diseases. *Rev. Neurosci.* **26**, 359–370. <https://doi.org/10.1515/revneuro-2014-0085> (2015).

Author contributions

S.K. did the molecular experiments. Z.J. performed part of the DNA modeling analyses. A.B. conceived and performed the DNA modeling and mathematics analyses, and wrote part of the manuscript. M.T. performed part of the mathematics analyses. H.A. collected the human samples and their clinical information. M.K. and R.M. provided critical comments and contributed to data collection. M.O. conceived and supervised the project, and wrote the manuscript.

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/s41598-021-99932-3>.

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

Reprints and permissions information is available at 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