2 Mechanism of conserved ancestral haplotype in SCA10

Karen McFarland[1], Vera Hashem[2], Anjana Tiwari[2], Linwei Zhang[3], Desmond Zeng[1], Jenny Holt[1], Justin Vincent[2], Maria Astudillo[2], Alex Cherches[2], Maria Aredondo[2], William G. Farmerie[4], Yu-Chih Tsai[5], Tyson A. Clark[6], She-Rui Gan[7], Tetsuya Takahashi[8], Ichiro Yabe[9], Birgitt Schüle[10], Astrid Rasmussen[11], Helio Teive[12], Tetsuo Ashizawa[2].

[1] McKnight Brain Institute, and Center for Translational Research in Neurodegeneration, University of Florida, Gainesville, FL, USA; [2] Houston Methodist Research Institute, Houston, TX, USA; [3] China Japan Friendship Hospital, Beijing, China; [4] Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, FL, USA; [5] PACBIO, Menlo Park, CA, USA; [6] GenapSys, Inc. Menlo Park, CA, USA; [7] Fujian University, Fuzhou, China; [8] Hiroshima University, Hiroshima, Japan; [9] Hokkaido University, Sapporo, Japan; [10] Stanford University School of Medicine, Stanford, CA, USA; [11] Oklahoma Medical Research Foundation, OK, USA; [12] Federal University of Parana, Curitiba, Brazil.

Presenting author: Tetsuo Ashizawa

Contact: tashizawa@houstonmethodist.org

Background: SCA10 is a rare disorder caused by an expanded intronic pentanucleotide repeat in *ATXN10*. SCA10 exclusively afflicts Latin Americans (LA) with Native American (NA) ancestry and East Asians (EA) with cerebellar ataxia and epilepsy. The expansion shows three configurations: (1) (ATTCT)n; (2) (ATTCT)n-(ATTCC)n, or (3) (ATTCT)n-(ATCCT)n-(ATCCC)n, each with different penetrance. Based on the SCA10 haplotype which is rare (<5%) in LA, NA and EA and <<0.1% elsewhere, we postulate that SCA10 expansion rose in EA on the rare haplotype and spread with prehistoric peopling of the Americas as a selection-neutral, low-penetrant mutation with occasional acquisition of pathogenic configurations.

Methods: Include (1) haplotyping, long-read SMRT/nanopore sequencing, and genotype-phenotype correlation using SCA10 DNAs, DNA from the 1000 Genomes Project (1KGP), and our own collection of general population DNAs from LA and EA, and (2) behavioral and histopathological analyses of transgenic animals expressing different repeat expansion configurations.

Results: All SCA10 families tested share the extended 91kb-haplotype that includes the G allele at rs41524745. This SNP is located 35kb centromeric to the SCA10 repeat (https://www.ncbi.nlm.nih.gov/snp/rs41524547#frequency_tab) juxtaposing the 5' end of the seed sequence of miR4762-5p. Absolute linkage was found between the G allele and the SCA10 repeat expansion despite the distance, which predicts multiple historical recombinations in this interval. Furthermore, 64 G(+) DNA from the 1KGP and 123 G(+) DNA of our 2,163 LA/EA general population samples showed that up to 25% of G(+) samples have SCA10 repeat expansions. In Peru, the SCA10 expansion may exist in 1% of the general population based on our 1KGP sample analysis! SCA10 expansions that have (ATTCC)n or (ATCCT)n(ATCCC)n repeat exhibit full penetrance while pure (ATTCT)n expansion has reduced penetrance in limited human and transgenic mouse data.

Discussions: Further studies, confirmatory data collections and experiments on miR4762-5p and recombination between rs41524547 and expanded SCA10 repeat, have begun under a new program of US-Brazil collaboration (supported with NIH 1R01NS115002-01; PIs: T. Ashizawa and L. Jardim). We wish to extend the project to other countries that have a population of SCA10 patients.