Proceedings of the 5th International Conference on Modern Medicine and Global Health
DOI: 10.54254/2753-8818/2027.38791

CAG Threshold and Genetic Modifiers Shape SCA3 Age at Onset

Mila Chen[1,*], Ruitong Zhang[2]

[1]Epsom College in Malaysia, Kuala Lumpur, Malaysia
[2]HD Shanghai School, Shanghai, China
**Corresponding Author. Email: mila.chen26@epsomcollege.edu.my*

Abstract. The abstract should summarize the content of the paper and should contain at least 150 and at most 250 words. The abstract should be a single paragraph. Do not include citations, figure numbers, table numbers, or displayed equations. Spinal cord cerebellar ataxia type 3 (SCA3), has significant differences in the individual onset of SCA3, which cannot be explained solely by abnormal repeat amplification of CAG. This paper compiles multiple studies related to the age of onset of SCA3, covering multicenter paired data from different ethnicities such as Europe, Brazil, and China. It clarifies the amplification quantity of the CAG gene in ATXN3 and the combined effect of genetic modification factors on the onset age. In summary, the age of onset of SCA3 proposed in this study is influenced by both the number of CAG repeats and genetic modification factors, and the modification factors themselves have racial specificity. At the clinical level, the combined detection of CAG repeat count and modifier factors can improve the accuracy of predicting the age of onset and facilitate the determination of more appropriate treatment plans. Future studies integrating multi-omics approaches and functional models may improve prediction accuracy and facilitate personalized therapeutic strategies for SCA3 patients.

Keywords: SCA3 age at onset, genetic modifiers, CAG threshold

1. Introduction

Spinocerebellar Ataxia Type 3 (SCA3), also called Machado-Joseph disease, ranks as the most common neurodegenerative disorder caused by expanded polyglutamine fragments. Different patients show a huge gap in the time of first symptoms.

Most affected people develop discomfort between twenty and fifty years old [1]. Typical signs include unsteady walking eye movement disorders and abnormal facial movements. Sleep troubles also appear in most cases. The disease comes from abnormal CAG repeats inside exon 10 of the ATXN3 gene located on chromosome 14q32.1. Mutated protein strands fold incorrectly and pile up inside nerve cells to trigger continuous nerve damage [2, 3].

Lots of large-scale patient data prove that longer CAG stretches in Ataxin-3 (ATXN3) link to earlier disease onset. The connection does not follow a simple straight line [4, 5]. Only fifty to seventy percent of age differences between patients can be explained by CAG length alone. Some people with identical ATXN3 repeats fall ill twenty years apart. This obvious gap brings our

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2027.38791

inspiration for further exploration. There are other genes with variable CAG sequences that can change the speed of disease progress. Thus the research question is: The way both ATXN3 CAG repeat length and genetic modifiers work together to change the age at onset of SCA3.

The most crucial problem is that SCA3 cannot be treated completely in a modern medical way. The most crucial problem is that SCA3 cannot be completely treated with current medical approaches. Therefore, genetic analysis can be used to investigate disease onset and progression, systematically characterize the independent regulatory patterns of CAG repeats, and summarize the interactions between mutant ATXN3 and different genetic modifiers. The overall analysis is expected to provide practical references for SCA3 genetic assessment and the development of targeted therapies. This paper aims to sort clinical genetic research papers focusing on the age at onset of SCA3 published at home and abroad. It objectively integrates research conclusions recorded in various documents. The coalition aims to summarize existing genetic findings of SCA3 onset age, sort out consistent laws and contradictory results across different ethnic cohorts, and further clarify research gaps for subsequent discussion.

2. Correlation between ATXN3 CAG repeats and age at onset

Patient follow-up records covering European, Brazilian, multi-national and Chinese cohorts all describe the association between the length of pathogenic CAG repeats in the ATXN3 gene and patients' initial symptomatic age. Three consistent core conclusions can be drawn after sorting these clinical records.

First, longer pathogenic CAG repeats in the ATXN3 gene correspond to an earlier appearance of initial symptoms such as unsteady walking and limb incoordination. Clinical follow-up data covering 7 homozygous and 30 heterozygous Chinese patients clearly reflect this inverse trend [4]. Clinical observations of 403 local Chinese patients also support the same rule. Patients carrying over 80 pathogenic repeats usually develop symptoms in adolescence, while those with repeats ranging from 55 to 60 mostly show obvious motor disorders around the age of fifty [2]. This trend is consistent in case records of all regions and ethnic groups.

Second, the regulatory effect of CAG repeats on disease onset has an obvious critical threshold instead of a simple linear change. Patient data collected from multiple Brazilian regions observe that minor fluctuations of repeat counts between 55 and 60 barely alter onset timing. Once repeats exceed 65 copies, neurotoxicity will be drastically amplified with each additional CAG unit [1]. From a molecular perspective, excessively long polyglutamine chains alter the spatial conformation of ataxin-3 proteins, which easily forms toxic deposits inside neurons and accelerates brain tissue damage. Summary data of 21 global homozygous patients show that dual expanded alleles on two chromosomes aggravate the tendency of early disease onset [3]. Apart from pathogenic alleles, the CAG length of wild-type ATXN3 also slightly influences onset time, and different combinations of normal and mutant alleles lead to varied clinical manifestations [6, 7]. Clinical data including 1439 genetically confirmed Chinese patients also record that longer ATXN3 pathogenic repeats generally correspond to younger ages of initial gait disturbance [6].

Third, the length of pathogenic CAG repeats alone cannot fully explain inter-individual differences in SCA onset age. Case records from all regions show that two patients carrying identical pathogenic CAG copies may have an onset age gap of over 20 years. Long-term follow-up of patients in the Azorean archipelago also reveals that numerous onset discrepancies remain unexplained even when CAG length is controlled as a fixed condition [5]. Among 7 Chinese homozygous cases, two subjects carried biallelic repeats of 55/56 and 56/58 respectively, yet their onset ages were 25 and 59 years old, creating a 34-year age gap [3]. Clinical summaries of 1212

DOI: 10.54254/2753-8818/2027.38791

symptomatic Chinese patients indicate that pathogenic CAG sequences alone fail to explain nearly 30% of individual phenotypic differences [6]. Follow-up data of 403 domestic patients also show that only pathogenic repeats can explain just over 60% of onset variance, and the inclusion of wild-type alleles only slightly improves the explanatory capacity. This inconsistency implies other genetic factors exist to regulate disease onset [7].

3. Genetic modifiers regulating the onset age of SCA3

Case studies of different ethnic groups uniformly demonstrate that single ATXN3 variants cannot account for all onset heterogeneity. Researchers across the globe have identified multiple genetic loci capable of altering disease progression. Integrated with cohort data from Europe and Brazil, clinical records of Chinese homozygous patients, inflammatory gene studies of Azorean subjects, large-scale Chinese TCF4 data and APOE follow-up documents, all verified modifiers are divided into five categories with their mutual pathological effects with mutant ataxin-3 summarized separately [4].

3.1. Polymorphic CAG modifiers identified in European and Brazilian cohorts

Joint multi-center follow-up records from Europe and nationwide SCA case summaries of Brazil state that genes with normal CAG polymorphisms including ATXN2, ATN1 and Huntingtin (HTT) exert regulatory impacts on SCA onset [1]. Medium-length CAG alleles of ATXN2 advance the emergence of motor symptoms. European carriers with 27–29 intermediate repeats generally develop disease earlier than those with short alleles, The in-depth analysis among 390 European SCA3 probands further expanded this understanding: in addition to the intermediate CAG repeat, the 9bp repeat of the ATXN2 gene is also significantly correlated with age of onset and disease progression rate, suggesting that the modification mechanism of ATXN2 on SCA3 may involve multiple levels such as splicing regulation and protein homeostasis , yet such variants only account for 2% to 5% of Brazilian patients, making this trend undetectable in local samples [8]. Long wild-type ATN1 alleles interact with mutant ataxin-3 to exacerbate neural injury and accelerate symptom presentation. In contrast, short wild-type HTT fragments carry mild protective effects that delay ataxia onset, which has also been documented in Azorean patient follow-up records [5].

Apart from polymorphic coding genes, biallelic expansion of ATXN3 represents a unique endogenous modifying factor. Summary clinical data of 21 global homozygous patients show an average onset age of 32.81 years, while 30heterozygous control subjects have a mean onset age of 49.90, presenting a notable inter-group difference [3].

3.2. Ethnic-specific intronic modifier: TCF4 in Chinese populations

Long-term clinical data covering 1439 genetically confirmed Chinese SCA3 patients identify intronic CAG repeats within the TCF4 gene as a novel modifier exclusively observed in Han Chinese cohorts [6].

Subjects are stratified into three groups based on the longer allele's repeat counts: Short (7–13 copies), Medium (14–39 copies), and Long (≥40 copies). Relevant documents note that TCF4 itself cannot directly determine onset age, but it amplifies the pathogenic potency of mutant ATXN3. Each additional pathogenic CAG copy shortens onset by 2.04 years in the Short group, 2.60 years in the Medium group and 3.18 years in the Long group. This modifying effect is most prominent in patients with initial symptoms before 25 years old [6]. Cellular research further elaborates the

DOI: 10.54254/2753-8818/2027.38791

bidirectional pathological cycle: mutant ataxin-3 aggravates splicing defects induced by long TCF4 intronic sequences, while expanded TCF4 repeats conversely promote intracellular aggregation of toxic ataxin-3 proteins, mutually exacerbating neuronal damage.

3.3. Controversial APOE modifier with contradictory ethnic results

Early small-scale European and American case reports claim that the APOE ε2 allele triggers earlier onset, while long-term follow-up data of 403 multi-region Chinese patients draw opposing conclusions [7, 8]. However, a case control research which includes 155 Han Chinese SCA3 patients has shown, APOE ε2 gene carriers would have an onset advance for 4 years, prompting that how APOE ε2 gene affects the onset completely depends on race [9]. This research completed genotyping of three mainstream APOE isoforms. After matching subjects with their respective ATXN3 repeat lengths, no significant onset age differences were found among patients with distinct APOE genotypes. Even when samples were stratified by gender or limited to subjects with disease duration less than five years, no APOE-related onset disparities could be observed. The conflicting outcomes between Western and Chinese cohorts are likely attributed to variant allele frequencies and varied linkage disequilibrium patterns among different ethnic populations.

3.4. Inflammatory SNP modifier loci

Genetic testing data of 86 Azorean patients reveal that promoter single-nucleotide variants of the IL6 gene act as another type of modifier independent of CAG sequences [4]. Patients carrying the c.-174C variant develop symptoms four years earlier on average compared with GG homozygotes. This variant reduces IL6 transcription in peripheral blood, disrupts cerebral inflammatory balance and accelerates neuronal loss. Previous papers once hypothesized synergistic pathogenic effects between this locus and APOE ε2, yet large-scale Chinese cohort data fail to validate such interaction. Other inflammatory genes such as IL1A, IL1B and TNF show no stable modifying capacity in recorded patient samples.

3.5. General summary of bidirectional synergistic pathogenic mechanisms

All verified modifiers interact with mutant ataxin-3 to produce synergistic neuronal damage via three pathways: abnormal protein aggregation, disrupted RNA transcription and chronic neuroinflammation. Long polymorphic CAG alleles, extended TCF4 intronic repeats and risk variants of IL6 accelerate disease onset, whereas short wild-type HTT fragments exert protective effects that delay clinical manifestations. Aberrant protein binding, defective gene splicing and sustained neuroinflammation constitute three core molecular mechanisms of genetic modification. The dosage effect caused by dual expanded ATXN3 alleles has been validated by clinical records of homozygous patients [3]. Beyond exogenous polymorphic genes, the matching pattern of wild-type and mutant ATXN3 alleles independently contributes to onset age variation [7]. Another study found in the SCA3/MJD cohort in southern Brazil that the rs3512 G/G genotype of the DNA repair gene FAN1 is independent of CAG length and ATXN2 and affects the age of onset. The combined analysis of FAN1 and ATXN2 can increase the explanatory power of age of onset variation from 55.2% to 63.0%. This discovery suggests that the DNA repair pathway may be another important regulatory layer after known modifying factors, and also indicates that multi-gene combination models have important value in improving the accuracy of predicting the age of onset [10].

DOI: 10.54254/2753-8818/2027.38791

4. Integrated comparison and research gaps

After organizing documents of all ethnic cohorts, horizontal comparisons and comprehensive inductions are conducted from three dimensions: universal laws, ethnic discrepancies and unaddressed research directions, to sort out the complete joint regulatory logic of two categories of genetic factors [5].

4.1. Consistent rules and ethnic differences

The negative correlation between ATXN3 CAG length and early onset is universally observed in European, Brazilian and East Asian patient cohorts. However, modifier allele frequencies vary greatly across ethnic groups.

Intermediate ATXN2 alleles are prevalent in European populations but rare in Brazilian samples; the regulatory function of TCF4 intronic repeats is only significant among Han Chinese; inflammatory variants exert stronger effects in isolated Azorean crowds; research conclusions regarding APOE remain contradictory between Western and Chinese cohorts [7]. The dosage modification phenomenon recorded in Chinese homozygous cases is rarely documented in European and American papers. Divergent genetic backgrounds and variant frequencies across populations lead to inconsistent screening results of modifier genes in regional studies.

4.2. Unified summary of complete regulatory logic

Expanded ATXN3 CAG repeats serve as the primary factor determining baseline neuronal damage severity, with damage rapidly aggravated once repeats exceed 65 copies. Based on this baseline, various modifiers either accelerate or slow down cerebral lesion progression through synergistic pathological interactions. Apart from exogenous polymorphic loci, the combination of wild-type and mutant ATXN3 alleles forms an intrinsic regulatory factor affecting onset age [7]. Accelerating factors include intermediate ATXN2 alleles, long TCF4 intronic repeats, IL6 risk variants and biallelic homozygous expansion, while short wild-type HTT alleles carry protective functions. The superposition of baseline neurotoxicity, gene dosage effects and multiple modifying signals results in the wide onset range of 5 to 70 years across global SCA3 patients [3].

4.3. Limitations of existing studies and future research directions

Summary of global case records shows that 24% to 34% of inter-individual onset differences still cannot be fully explained even when combining ATXN3 repeats and all confirmed modifiers. In addition, in recent years, research has begun to focus on the modifying effect of protein degradation pathways on the age of onset of SCA3. Martins et al found in 287 SCA3/MJD patients that the G allele at position r s1559085 of the CAST gene is independent of CAG repeat length and affects the age of onset, suggesting that the calpain cleavage pathway may be the fourth modification pathway after protein aggregation, RNA toxicity, and neuroinflammation [11]. Current literature lacks systematic exploration of rare copy number variants, epigenetic regulation such as ATXN3 promoter methylation and multi-gene synergistic effects. Controversial loci like APOE require verification through larger multi-ethnic cohorts [7]. Most existing clinical research only relies on population observational data, lacking validation via cellular and animal models of modifying mechanisms, and cross-ethnic follow-up studies focusing on homozygous dosage effects remain insufficient [6].

DOI: 10.54254/2753-8818/2027.38791

5. Conclusion

This study focuses on how the threshold effect of ATXN3 gene CAG repeat growth and the regulatory mechanism of modifier factors jointly affect the age of onset of SCA3. Three main conclusions were obtained: CAG repeat length is the main determining factor of SCA3 age of onset. Through large-scale comparative analysis, this study verified that CAG repeat is negatively correlated with age of onset. However, the threshold effect can only serve as a necessary condition for advancing the age of onset of the disease. The significant dispersion of patients' clinical phenotypes under a uniform number of repetitions explains that there are still other factors that affect the age of onset, such as genetic modification factors. The modifying factors form a synergistic regulatory system through three pathways - protein, RNA, and inflammation- and this system has significant race specificity: This paper analyzed multiple data and found that each of the four pathways independently acts on the SCA3 disease modification effect. In population genetics, it has been shown that there are significant differences in the direction of effects of most modification sites between races. TCF4, as a species-specific modifier which is unique to the Chinese Han SCA3 population, has a stronger effect than other common modification sites. The TCF4 gene can cause changes in the onset time of the disease in the Han population, and this effect is not apparent in other ethnic groups, highlighting its racial specificity. Future research should integrate multi-omics analysis, large-scale multi-ethnic cohorts, and cellular or animal models to further elucidate the synergistic regulatory mechanisms by which factors such as genetic variants, epigenetic regulation, and protein degradation pathways influence the variation in age of onset for SCA3.

Authors contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] de Castilhos, R. M., Furtado, G. V., Gheno, T. C., Schaeffer, P., Russo, A., Barsottini, O., Pedroso, J. L., Salarini, D. Z., Vargas, F. R., Lima, M. A. D., Godeiro, C., Santana-da-Silva, L. C., Toralles, M. B., Santos, S., van der Linden, H., Wanderley, H. Y., Vanconcelos de Medeiros, P. F., Pereira, E., Ribeiro, E., Saraiva-Pereira, M. L., & Jardim, L. B. (2014). Spinocerebellar ataxias in Brazil—Frequencies and modulating effects of related genes. *Cerebellum, 13*(1), 17–28. https://doi.org/10.1007/s12311-013-0510-y

[2] Chen, Z., Zheng, C., Long, Z., Cao, L., Li, X., Shang, H., Yin, X., Zhang, B., Liu, J., Ding, D., Peng, Y., Peng, H., Ye, W., Qiu, R., Pan, Q., Xia, K., Chen, S., Sequeiros, J., Ashizawa, T.,... Jiang, H. (2016). (CAG)n loci as genetic modifiers of age-at-onset in patients with Machado-Joseph disease from mainland China. *Brain, 139*(8), e41. https://doi.org/10.1093/brain/aww087

[3] Li, Q. F., Cheng, H. L., Zhao, J. J., Dong, Y., Yang, L., Ma, Y., & Wu, Z. Y. Clinical features and genetic characteristics of homozygous spinocerebellar ataxia type 3. *Molecular Genetics & Genomic Medicine, 8*(6), e1314. https://doi.org/10.1002/mgg3.1314

[4] Raposo, M., Bettencourt, C., Ramos, A., Kazachkova, N., Vasconcelos, J., Kay, T., Bruges-Armas, J., & Lima, M. (2016). Promoter variation and expression levels of inflammatory genes IL1A, IL1B, IL6 and TNF in blood of spinocerebellar ataxia type 3 patients.

[5] Tezenas du Montcel, S., Durr, A., Bauer, P., Figueroa, K. P., Ichikawa, Y., Brussino, A., Forlani, S., et al. (2014). Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain, 137*(9), 2444–2455. https://doi.org/10.1093/brain/awu174

[6] Wan, N., Wu, Q., Wan, L., He, L., Shi, Y., Qiu, R., Jiang, Q., Peng, L., Wang, C., Lei, L., Long, Z., Tang, B., Chen, Z., & Jiang, H. (2026). TCF4 intronic CAG repeat length modulates the effect of ATXN3 on age at onset in spinocerebellar ataxia type 3. *Journal of Advanced Research*. https://doi.org/10.1016/j.jare.2026.02.031

[7] Zhou, Q., Ni, W., Dong, Y., Wang, N., Gan, S. R., & Wu, Z. Y. (2014). The role of apolipoprotein E as a risk factor for an earlier age at onset for Machado-Joseph disease. *PLOS ONE, 9*(11), e111356.

DOI: 10.54254/2753-8818/2027.38791

https://doi.org/10.1371/journal.pone.0111356

[8] Lauerer, M., et al. (2025). ATXN2 intermediate CAG repeats and 9-bp repeat influences onset and progression in spinocerebellar ataxia type3. *Acta Neuropathological Communications*. Advance online publication. https://doi.org/1186/s40478-025-02015-x

[9] Peng, H., Yang, M., Chen, Z., et al. (2014). APOE & 2 allele may decrease the age at onset in patients with spinocerebellar ataxia type 3/Machado-Joseph disease. *Neurobiology of Aging, 35*(9), 2179.15-2179.18. https://doi.org/10.1016/j.neurobiolaging.2014.03.020

[10] Mergener, R., et al. (2020). FAN1 variants and ATXN2 CAG repeat length modulate age at onset in spinocerebellar ataxia type 3/Machado-Joseph disease. *Neuromolecular Medicine, 22*(1), 133-138. https://doi.org/10.1007/s12017-019-08572-4

[11] Martins, A. C., et al. (2021). Calpain system genes are associated with age at onset in spinocerebellar ataxia type 3/Machado-Joseph disease. *Journal of Molecular Neuroscience, 71*(9), 1906-1913. https://doi.org/10.1007/S12031-021-01877-9