

Frequency of the loss of CAA interruption in the *HTT* CAG tract and implications for Huntington disease in the reduced penetrance range

Hailey Findlay Black, BSc¹, Galen E. B. Wright, PhD¹, Jennifer A. Collins, BA¹, Nicholas Caron, PhD¹, Chris Kay, PhD¹, Qingwen Xia, MD¹, Larissa Arning, PhD², Emilia K. Bijlsma, MD, PhD³, Ferdinando Squitieri, MD, PhD⁴, Huu Phuc Nguyen, MD² and Michael R. Hayden, MB, ChB, PhD¹

Purpose: In some Huntington disease (HD) patients, the “loss of interruption” (LOI) variant eliminates an interrupting codon in the *HTT* CAG-repeat tract, which causes earlier age of onset (AOO). The magnitude of this effect is uncertain, since previous studies included few LOI carriers, and the variant also causes CAG size misestimation. We developed a rapid LOI detection screen, enabling unbiased frequency estimation among manifest HD patients. Additionally, we combined published data with clinical data from newly identified patients to accurately characterize the LOI’s effect on AOO.

Methods: We developed a LOI detection polymerase chain reaction (PCR) assay, and screened patients to estimate the frequency of the LOI variant and its effect on AOO.

Results: Mean onset for LOI carriers ($n = 49$) is 20.4 years earlier than expected based on diagnosed CAG size. After correcting for

CAG size underestimation, the variant is still associated with onset 9.5 years earlier. The LOI is present in 1.02% of symptomatic HD patients, and in 32.2% of symptomatic reduced penetrance (RP) range patients (36–39 CAGs).

Conclusion: The LOI causes significantly earlier onset, greater than expected by CAG length, particularly in persons with 36–39 CAG repeats. Detection of this variant has implications for HD families, especially for those in the RP range.

Genetics in Medicine (2020) 22:2108–2113; <https://doi.org/10.1038/s41436-020-0917-z>

Keywords: Huntington disease; age of onset; genetic modifiers; reduced penetrance; repeat expansion

INTRODUCTION

Huntington disease (HD, OMIM 143100) is an autosomal-dominant neurodegenerative disorder caused by CAG-repeat expansion within the *HTT* gene. HD is characterized by progressive cognitive, psychiatric, and motor symptoms.¹ Initial motor signs define clinical disease onset,^{2,3} usually occurring in midadulthood.

Polymerase chain reaction (PCR)-based fragment sizing of CAG repeats is used for genetic confirmation in symptomatic individuals, and presymptomatic predictive testing in at-risk individuals.¹ *HTT* alleles with ≤ 35 CAG repeats are usually⁴ nonpathogenic, but intermediate alleles (27–35 repeats) may undergo germline expansion to pathogenic sizes.^{5,6} Alleles with ≥ 40 CAG repeats are pathogenic within normal lifespans.

A recent study of reduced penetrance (RP) allele frequency in the general population estimated penetrance of 36–39 CAG-repeat alleles at $<1\%$ by age 65,⁷ markedly lower than the 63.9% penetrance by age 65 for repeat sizes 36–39 within clinically ascertained families.⁸

CAG length, the major determinant of age of onset (AOO),^{9–11} accounts for approximately 70% of variation of AOO.^{9–11} AOO variance increases with decreased CAG length and most individuals with CAG sizes of 36–39 (the RP range) will not manifest with HD within a normal lifespan.^{10,12}

Recently, a known variant¹³ LRG_763t1:c.118A>G, the loss of the CAA interruption penultimate to the 3' end of the *HTT* CAG tract (Fig. 1a), was identified as a *cis*-acting modifier of AOO.^{14–16} This loss of interruption (LOI) variant is associated with significantly earlier AOO when present on pathogenic range HD alleles,^{14–16} with consistently early onset observed within carrier families.¹⁴ However, since previous studies included relatively few patients (i.e., $n \leq 21$), the magnitude of this effect on AOO has remained uncertain.

Importantly, it remains unclear whether the LOI variant’s effect on AOO is a true biological effect, or merely an artifact of diagnostic underestimation of the uninterrupted CAG length. The variant causes the loss of the interrupting CAA

¹Centre for Molecular Medicine and Therapeutics, Department of Medical Genetics, University of British Columbia, Vancouver, Canada; ²Department of Human Genetics, Medical Faculty, Ruhr University of Bochum, Bochum, Germany; ³Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; ⁴Huntington and Rare Diseases Unit, Fondazione IRCCS Casa Sollievo della Sofferenza Research Hospital, San Giovanni Rotondo, Italy. Correspondence: Michael R. Hayden (mrh@cmmt.ubc.ca)

Submitted 29 May 2020; revised 17 July 2020; accepted: 17 July 2020

Published online: 3 August 2020

codon but does not cause diagnostic sizing failure, as standard primers still recognize the variant sequence. Consequently, although polyglutamine repeats are counted accurately, uninterrupted CAG tracts are underestimated by two trinucleotide repeats for LOI carriers, since CAG-length inference assumes canonical interrupting sequences (Fig. 1b). Since uninterrupted CAG repeats predict AOO better than polyglutamine repeats,^{14,15} it has been unclear to what extent this CAG-length underestimation fully underlies the earlier onset in LOI carriers, or whether in addition there is a biological effect beyond this underestimate.

We developed a novel PCR assay to identify LOI carriers, and screened DNA from all unrelated HD pedigrees in the University of British Columbia (UBC) HD Biobank, enabling unbiased estimation of variant frequency among affected patients. Additionally, we have combined clinical data from additional patients identified through our screen with previously published data to better characterize the LOI variant's effect on AOO. This represents a 2.3-fold increase in LOI carrier sample size over any published cohort.

In this report, we identify a true biological effect of the LOI on AOO beyond increasing uninterrupted CAG length. We demonstrate that the implications of the LOI are particularly relevant to persons with RP range CAG-repeat lengths. We therefore propose that this variant should be tested for and considered when interpreting predictive testing results, particularly for those in the RP range.

MATERIALS AND METHODS

Ethics statement

Samples and clinical data were collected with informed consent and ethical approval from the UBC Children's and Women's Health Centre Research Ethics Board (H06-70467, H06-70410) and the Ruhr University Bochum Medical Faculty Ethics Committee (18-6563-BR).

PCR assay for LOI variant detection

We optimized a two-step PCR assay for LOI detection. First, we amplified a CAG tract-surrounding region using Roche Taq polymerase (11146165001) using the following primers: forward 5'-GGGACGCAAGGCGCCGTAG; reverse 5'-GGCTGAGGAAGCTGAGGAG. Cycle conditions were as follows: 7 minutes denaturation (95 °C), 35 cycles (30 seconds at 95 °C, 30 seconds at 63 °C, 30 seconds at 72 °C), 7 minutes extension (72 °C). The forward primer contains a mismatch allowing selective amplification of the rs13102260 major allele (G), which reduces off-target amplification downstream.

We diluted these PCR products 1:10, then performed PCR using AmpliTaq Gold360 DNA Polymerase (4398833), with 1 uL of diluted product per 25 uL PCR reaction. The forward primer sequence for this step (5'-GCAGCAGCAGCAGCCG CCG) is LOI-specific; reverse primer and cycle conditions were identical to the previous step.

We separated products on 3% agarose and confirmed positive samples by clonal Sanger sequencing as described.¹⁴

Cohort information

To determine LOI variant frequency in the symptomatic HD population, we assembled and screened genomic DNA samples from 491 unrelated affected HD probands. This cohort is comprised of DNA from the first affected family member with known CAG size and AOO from each UBC HD Biobank pedigree.

To estimate LOI frequency among symptomatic individuals in the RP range, we identified all affected individuals with reported allele sizes 36–39 from UBC ($n = 37$) and Bochum ($n = 22$) biobanks, and screened these 59 RP subjects for the LOI.

To determine the effect of the LOI on AOO, we compiled data from all available published LOI carriers ($n = 16$ individuals reported by Wright *et al.*;¹⁴ $n = 21$ individuals reported by GeM-HD¹⁵) and individuals newly identified in this study ($n = 12$), for a total of 49 individuals. GeM-HD CAG lengths and AOO for LOI carriers were extracted from published information using WebPlotDigitize (<https://automeris.io/WebPlotDigitizer/>) based on harmonized results of three independent raters.

For comparison, we identified 473 symptomatic non-LOI carriers with known CAG sizes and AOO; this group excluded individuals with known sequence variants in the interrupting region, and individuals with CAG sizes >52 to match the upper CAG size of the LOI samples.

AOO statistical analyses

We predicted mean AOO for each CAG size according to Langbehn *et al.*¹⁰ and compared with observed AOO in years earlier (predicted onset–observed onset) and as a ratio (observed onset ÷ predicted onset).

Because diagnostic sizing underestimates true uninterrupted CAG size in LOI carriers, we analyzed predicted AOO for LOI carriers using both diagnostic CAG size (from fragment analysis) and true uninterrupted CAG size (diagnostic CAG size + 2).

We performed statistical analyses using R. We identified significant differences in AOO between genotypes using Wilcoxon rank-sum tests, and fitted a mean predicted AOO curve to the LOI data using an exponential model.

RESULTS

The LOI variant is present in ~1% of symptomatic HD patients

Of 491 unrelated symptomatic patients (CAG 37–88), we identified 5 LOI carriers, putting the variant frequency at 1.02% (95% CI: 0.13–1.91%) within the symptomatic population. In all cases, the LOI was on the expanded allele. This cohort includes all unrelated probands in the UBC Biobank with available CAG size and AOO data.

LOI carriers exhibit significantly earlier onset across all uninterrupted CAG sizes

We compared AOO in 49 individuals carrying the LOI variant to predicted mean AOO using the Langbehn formula.¹⁰ Based

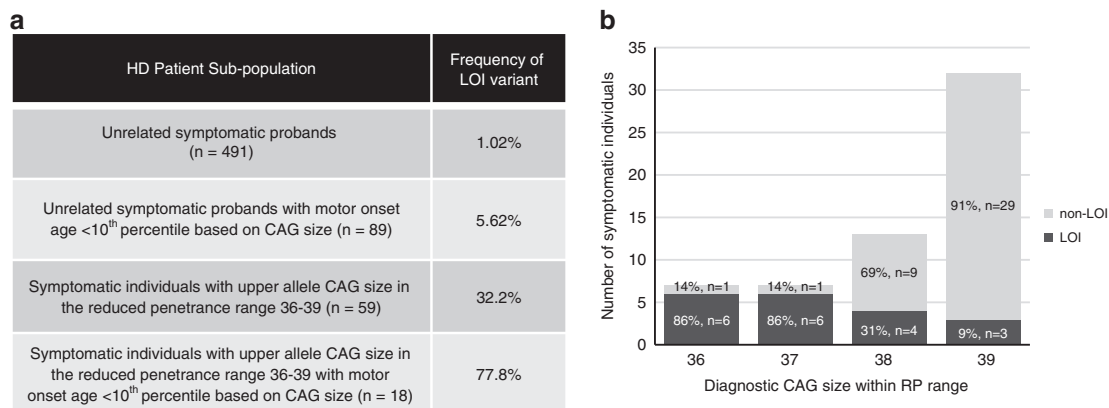


Fig. 2 Prevalence of the LOI variant in HD patient sub-populations. (a) Prevalence of the loss of interruption (LOI) variant among symptomatic Huntington disease (HD) populations as a proportion of total patients carrying the LOI allele. Prevalence is presented for the overall HD symptomatic population, among cases with onset earlier than predicted for diagnostic CAG size (<10th percentile), within the reduced penetrance (RP) allele range (36–39), and among cases in the RP allele range with onset earlier than predicted for CAG size (<10th percentile). (b) Proportion of patients carrying the LOI at each CAG size within the RP range (diagnostic CAG sizes 36–39).

on predicted onset according to diagnostic CAG sizing, LOI carriers present on average 20.4 years earlier than expected. (Fig. 1c). After correcting for CAG size underestimation to reflect true uninterrupted CAG sizes, LOI carriers still showed a mean of 9.5 years earlier onset. (Fig. 1c). LOI carriers showed significantly earlier AOO based on either diagnostic ($p = 2.2 \times 10^{-16}$) or uninterrupted CAG size ($p = 2.8 \times 10^{-12}$).

To eliminate biases caused by averaging across individuals with different CAG sizes, we determined AOO ratios for all individuals (actual onset age ÷ expected onset age). AOO ratio significantly differed between canonical repeat interruption carriers (AOO ratio = 0.96) and LOI carriers, based either on diagnostic CAG size ($0.68, p = 2.2 \times 10^{-16}$) or uninterrupted CAG size ($0.81, p = 3.1 \times 10^{-10}$) (Fig. 1d).

To determine whether CAG size-dependent effects drive hastened onset, we plotted LOI carriers by CAG size (Fig. 1e). Mean AOO is earlier than Langbehn-predicted onset based on either diagnostic or uninterrupted CAG size across all repeat sizes examined (36–50).

The LOI is enriched among early onset and symptomatic RP range cases

Individuals in our screening cohort with AOO within the 10th percentile for their CAG size showed 5.5-fold enrichment for the LOI (Fig. 2a).

We also assembled a cohort of symptomatic RP allele carriers, comprising all symptomatic individuals with diagnostic CAG-repeat lengths of 36–39 ($n = 59$). This cohort was markedly enriched for the LOI, which was present in 32.2% of these symptomatic RP allele carriers, and 86% of those with diagnostic CAG sizes of 36 or 37 (Fig. 2a, b). Among symptomatic RP allele carriers with AOO within the 10th percentile for their CAG size, 77.8% had the LOI (Fig. 2a).

DISCUSSION

Diagnostic CAG sizing underestimates true uninterrupted CAG size for LOI carriers

Individuals at risk for HD who request predictive testing frequently cite reasons involving making informed choices about their future, including education or careers.¹⁷ These decisions are likely influenced by information these individuals receive regarding their likelihood of presenting symptomatically with HD.

It is crucial to recognize that uninterrupted CAG tract lengths are underestimated by two CAG repeats in LOI carriers (Fig. 1b). For LOI carriers, information on potential AOO will be more inaccurate for individuals with shorter CAG tracts, in whom two additional uninterrupted repeats constitute a larger percentage of total CAG count. In some cases, underestimation by two CAG repeats means LOI carriers actually have 40 or 41 repeats, but are mistakenly informed they fall within the RP range. In our cohort, 15.6% of symptomatic patients with diagnostic CAG sizes 38–39 were inaccurately informed that they carry an RP allele.

These inaccuracies have consequences beyond assessment of onset, including eligibility for clinical trials enrolling individuals within specific CAG size or CAG-age product score ranges.

LOI presence may influence penetrance of RP range alleles

Symptomatic RP range individuals have significant enrichment for the LOI variant (~30-fold more frequent than in fully penetrant alleles). The LOI is found in the vast majority of symptomatic RP range individuals with CAG sizes 36–37 (86%), or with AOO in the earliest 10th percentile for their CAG size (77.8%), suggesting that the LOI may influence whether RP range individuals develop HD during their expected lifetime. We examined all available premanifest RP subjects in the UBC Biobank ($n = 59$), but could not assess

how long RP individuals lacking the LOI remain asymptomatic, because we lacked an adequate cohort of sufficiently old asymptomatic RP allele carriers. We hypothesize that RP range individuals with the LOI more frequently become symptomatic within a normal lifespan. This may explain why general population RP allele penetrance⁷ is lower than penetrance within clinically ascertained families.⁸

LOI screening during HD predictive testing could improve patient information

Screening for the LOI during predictive testing for HD would improve accuracy of CAG sizing results and relevance of information patients are given, particularly for persons with 36–39 repeats. As shown in Fig. 1c–e, correcting diagnostic sizing of LOI carriers to reflect true uninterrupted CAG sizes does not fully eliminate the discrepancy between actual and predicted onset in these patients. Individuals with a family history of RP alleles or earlier-than-expected onset are a particularly high priority for LOI screening.

Correcting CAG size reduces, but does not eliminate, the LOI variant's effect on AOO, suggesting a CAG-independent biological effect

Correcting for CAG-repeat underestimation in LOI carriers does not eliminate the effect of the LOI on earlier AOO. Although the mechanism remains to be determined, we have previously proposed that the increased somatic instability observed in LOI alleles compared with alleles with canonical repeat interruptions may produce huntingtin protein with even longer polyglutamine lengths, causing greater damage and earlier onset.¹⁴ Repeat instability has been implicated as a modifier of numerous repeat expansion disorders, including HD,^{16,18} and is observed to increase in the absence of repeat interruptions,¹⁹ making this mechanism plausible. However, it is possible that the LOI acts through other mechanisms, such as translation efficiency alteration or increased repeat associated non-AUG (RAN) translation toxicity.²⁰

In conclusion, three independent groups^{14–16} have observed that the LOI hastens AOO in small cohorts. Now, our combined and updated data set has elucidated LOI frequency and refined the magnitude of the effect on AOO. Finally, by developing and optimizing a novel LOI-specific PCR assay, we provide a reliable tool to offer routine screening to patients who could be impacted by inheritance of the LOI.

SUPPLEMENTARY INFORMATION

The online version of this article (<https://doi.org/10.1038/s41436-020-0917-z>) contains supplementary material, which is available to authorized users.

ACKNOWLEDGEMENTS

This work was supported by a CIHR Foundation Grant to M.R.H. (FDN-154278), an E-Rare Research Program: TreatPolyQ grant (BMBF 01GM1804) to M.R.H. and H.P.N., a CIHR CGS-M Scholarship (H.F.B.), the James Family and HDSA Berman/Topper Fellowships (N.C.), and the BC Children's Hospital Research

institute. We thank the HD patients and families who have participated in research through sample and data contributions to the UBC HD BioBank, Ruhr-Universität Bochum, Germany, and the LIRH Foundation and Mendel Institute, Rome, Italy.

DISCLOSURE

H.P.N. consulted for Astra Zeneca and Skyhawk Therapeutics. M. R.H. sits on the Boards of Directors of Xenon, Aurinia, Ionis, and 89 Bio, for which he receives directors' compensation. F.S. provided consulting services and advisory board functions to Teva, Wave Pharma, Pfizer, Hoffmann-La Roche, Novartis, PTC Therapeutics, UCB, Oman Ministry of Health–Sultanate of Oman, and is a cofounder, scientific officer, and consultant of the Italian League for Research on Huntington and related diseases (LIRH Foundation). The other authors declare no conflicts of interest.

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

REFERENCES

- Caron NS, Wright GEB, Hayden MR. Huntington disease. 1998 Oct 23 [Updated 2018 Jul 5]. In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews. Seattle, WA: University of Washington, Seattle; 1993–2020.
- Testa CM, Jankovic J. Huntington disease: a quarter century of progress since the gene discovery. *J Neurol Sci.* 2019;396:52–68.
- Ross CA, Reilmann R, Cardoso F, et al. Movement Disorder Society Task Force viewpoint: Huntington's disease diagnostic categories. *Mov Disord Clin Pract.* 2019;6:541–546.
- Squitieri F, Jankovic J. Huntington's disease: how intermediate are intermediate repeat lengths? *Mov Disord.* 2012;27:1714–1717.
- Semaka A, Creighton S, Warby S, Hayden MR. Predictive testing for Huntington disease: interpretation and significance of intermediate alleles. *Clin Genet.* 2006;70:283–294.
- Kay C, Collins JA, Wright GE, et al. The molecular epidemiology of Huntington disease is related to intermediate allele frequency and haplotype in the general population. *Am J Med Genet B Neuropsychiatr Genet.* 2018;177:346–357.
- Kay C, Collins JA, Miedzybrodzka, et al. Huntington disease reduced penetrance alleles occur at high frequency in the general population. *Neurology.* 2016;87:282–288.
- Quarrell OW, Rigby AS, Barron L, et al. Reduced penetrance alleles for Huntington's disease: a multi-centre direct observational study. *J Med Genet.* 2007;44:e68.
- Rosenblatt A, Brinkman RR, Liang KY, et al. Familial influence on age of onset among siblings with Huntington disease. *Am J Med Genet.* 2001;105:399–403.
- Langbehn DR, Brinkman RR, Falush D, Paulsen JS, Hayden MR. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. *Clin Genet.* 2004;65:267–277.
- Wexler NS. Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proc Natl Acad Sci USA.* 2004;101:3498–3503.
- Rubinsztein DC, Leggo J, Coles R, et al. Phenotypic characterization of individuals with 30–40 CAG repeats in the Huntington disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36–39 repeats. *Am J Hum Genet.* 1996;59:16.
- Goldberg YP, McMurray CT, Zeisler J, et al. Increased instability of intermediate alleles in families with sporadic Huntington disease compared to similar sized intermediate alleles in the general population. *Hum Mol Genet.* 1995;4:1911–1918.
- Wright GE, Collins JA, Kay C, et al. Length of uninterrupted CAG, independent of polyglutamine size, results in increased somatic instability, hastening onset of Huntington disease. *Am J Hum Genet.* 2019;104:1116–1126.

15. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. CAG repeat not polyglutamine length determines timing of Huntington's disease onset. *Cell*. 2019;178:887–900.
16. Ciosi M, Maxwell A, Cumming SA, et al. A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes. *EBioMedicine*. 2019;48:568–580.
17. Holman MA, Quillin J, York TP, Testa CM, Rosen AR, Norris VW. The changing age of individuals seeking presymptomatic genetic testing for Huntington disease. *J Genet Couns*. 2018;27:1157–1166.
18. Schmidt MH, Pearson CE. Disease-associated repeat instability and mismatch repair. *DNA Repair*. 2016;38:117–126.
19. Nolin SL, Glicksman A, Tortora N, et al. Expansions and contractions of the FMR1 CGG repeat in 5,508 transmissions of normal, intermediate, and premutation alleles. *Am J Med Genet*. 2019;179:1148–1156.
20. Zu T, Gibbens B, Doty NS, et al. Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci USA*. 2011;108:260–265.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, which permits any non-commercial 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 license, and indicate if changes were made. If you remix, transform, or build upon this article or a part thereof, you must distribute your contributions under the same license as the original. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>.

© The Author(s) 2020