

natural selection. If *cis*-elements of this sort did exist, it would therefore be most likely that they would be observed on haplotype backgrounds with a smaller-than-average number of repeats, even if they were associated with an average or higher-than-average disease incidence. This pattern is not consistent with the repeat length distribution for the A haplogroup or its variants. A *cis*-element with a specific effect increasing the mutation rate specifically for shorter alleles is more consistent with the data, because for example it can generate the bimodal distribution of repeat lengths within the general population observed for haplogroup A. However, invoking an element that induces peculiar mutational properties seems unparsimonious and does not on its own account for the difference in CAG length and disease incidence among the variants of the A haplogroup.

Daniel Falush<sup>1,\*</sup>

<sup>1</sup>Department of Microbiology, University College Cork, Cork, Ireland

\*Correspondence: [d.falush@ucc.ie](mailto:d.falush@ucc.ie)

### Acknowledgments

S. Warby, S. Butland, and two anonymous reviewers provided numerous helpful comments. The author is funded by Science Foundation of Ireland grant 05/FE1/B882.

### Web Resources

The URL for data presented herein is as follows:

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/OMIM/>

### References

1. Warby, S.C., Montpetit, A., Hayden, A.R., Carroll, J.B., Butland, S.L., Visscher, H., Collins, J.A., Semaka, A., Hudson, T.J., and Hayden, M.R. (2009). CAG expansion in the Huntington disease gene is associated with a specific and targetable predisposing haplogroup. *Am. J. Hum. Genet.* 84, 351–366.
2. Falush, D., Almquist, E.W., Brinkmann, R.R., Iwasa, Y., and Hayden, M.R. (2001). Measurement of mutational flow implies both a high new-mutation rate for Huntington disease and substantial underascertainment of late-onset cases. *Am. J. Hum. Genet.* 68, 373–385.
3. Libby, R.T., Hagerman, K.A., Pineda, V.V., Lau, R., Cho, D.H., Baccam, S.L., Axford, M.M., Cleary, J.D., Moore, J.M., Sopher, B.L., et al. (2008). CTCF *cis*-regulates trinucleotide repeat instability in an epigenetic manner: A novel basis for mutational hot spot determination. *PLoS Genet.* 4, e1000257.
4. Rubinsztein, D.C., Amos, W., Leggo, J., Goodburn, S., Ramesar, R.S., Old, J., Bontrop, R., McMahon, R., Barton, D.E., and Ferguson-Smith, M.A. (1994). Mutational bias provides a model for the evolution of Huntington's disease and predicts a general increase in disease prevalence. *Nat. Genet.* 7, 525–530.
5. Paradisi, I., Hernández, A., and Arias, S. (2008). Huntington disease mutation in Venezuela: Age of onset, haplotype analyses and geographic aggregation. *J. Hum. Genet.* 53, 127–135.

DOI 10.1016/j.ajhg.2009.11.006. ©2009 by The American Society of Human Genetics. All rights reserved.

---

## Response to Falush: A Role for *cis*-Element Polymorphisms in HD

*To the Editor:* We thank Falush for his important insights into the evolution of CAG expansion in the huntingtin (*HTT*) gene. Following observations from his computational model, Falush argues that the patterns observed in the genotyping at the *HTT* locus<sup>1</sup> can be explained by a mutational mechanism that is solely dependent on the size of the CAG tract, and that the evolution of Huntington's disease (HD) chromosomes is most simply explained by a common founder CAG expansion. On this basis, he argues that there is no need to invoke *cis*-elements having a role in the evolution of HD chromosomes.

We agree that founder CAG expansions likely play a role in the evolution of HD chromosomes in European populations. However, there remain several observations in the genotyping data that are difficult to reconcile with the hypothesis that HD chromosomes evolved exclusively from a common CAG-expanded founder. Furthermore,

invalid assumptions made in Falush's computational model weaken the argument for this hypothesis. Instead, we argue that the simplest explanation for our data is that *cis*-elements make an important contribution to CAG instability at the *HTT* locus. We propose that *cis*-element polymorphisms are an influential force behind the apparent multiple founder chromosomes based on the observed pattern of haplotypes in the European population and strong biological precedents for *cis*-elements influencing trinucleotide instability.

In our original publication, we described three important patterns in the genotyping data.<sup>1</sup> The first observation was that CAG expansion in the European population was highly enriched in very specific haplogroup A variants (A1, A2, and A3), but not in other haplogroup A variants (A4 and A5) or other haplogroups (B and C). It is important to note that the two variants with the strongest disease association (A1 and A2) are less similar to each other than either variant is to other non-disease-associated variants; the A1 variant is more closely related to A4 and A5 (differs at one or two SNP positions from A1) than to A2 (differs at three SNP positions). This observation makes it less likely that A1 and A2 are derived from a simple

common CAG-expanded founder and more likely that CAG expansion arose independently in these haplotypes.

The computational model described by Falush argues that the multiple founders are the result of genetic drift. This model assumes that CAG instability is strictly CAG-size dependent and that small random increases in the modal CAG size of these specific haplotypes (Falush proposes 17 to >22 CAG in A1, 17 to 20 CAG in A2, and is unclear for A3) were sufficient to cause an enrichment of HD mutations in these haplotypes that are otherwise neutral. However, if haplotypes (*cis*-elements) play no role in CAG instability, it is not clear why genetic drift would favor expansion multiple times in similar variants of haplogroup A rather than in the other common haplogroup in the general population (haplogroup C). The genetic drift mechanism proposed by Falush appears to be nonrandom. Haplogroup C also has chromosomes in the 20–22 CAG range, but very few expanded into the HD range (we detected none with >35 CAG in our sample). Additionally, CAG expansion is much more common in variant A3 compared to A4, despite having the same modal CAG size of 17 repeats (Figure 4C in Warby et al.<sup>1</sup>). These discrepancies make it difficult to explain the enrichment of CAG expansion in specific haplotypes through simple genetic drift and CAG dependency alone.

The second observation in our genotyping data was that the pattern of SNP linkage was punctuated rather than decaying as a function of genetic distance, as would be expected for a recent founder chromosome. We agree with Falush that this punctuated pattern could be explained by secondary changes such as new SNPs or gene conversion events, given enough time.

The third observation was that we did not find haplotypes that were unique to expanded chromosomes. We had anticipated that secondary genetic events, such as those generating the punctuated pattern of SNP linkage described above, would produce haplotypes that were found exclusively on CAG-expanded chromosomes. However, even the highest-risk haplotypes (A1 and A2) had CAG tract sizes that extended down into the normal (<27 CAG) range. If the CAG-expanded founder and nonexpanded chromosomes have been separated for substantial time, why has the founder not undergone further genetic variation that discriminates between them?

One possibility is that the haplotypes with founder CAG expansions (and all variants of that founder) have contracted back to the normal CAG size range, as explained by the computational model. We argue that this is unlikely because of the long timescale required for these types of change (see below). The other possible explanation is that the multiple founder CAG expansions are always very small expansions in CAG size, such as those that could result from genetic drift. In addition to the inconsistencies with the genetic drift model stated above, the hypothesis that modest increase (17 to 22) in modal

CAG size can result in a massive enrichment of CAG expansion in that haplotype relies heavily on insights from the computational model. However, the validity of the model depends on a few key assumptions that require further examination.

The first assumption of Falush's computational model is that for all CAG tract sizes, the mutation rate is CAG dependent such that it is proportional to the CAG tract size to the power of eight. This mutation rate is based on sperm typing data from 26 men in a Venezuelan HD cohort with CAG sizes ranging from 37 to 62 repeats.<sup>2</sup> We agree that this mutational model appropriately describes changes to the CAG tract in sperm for chromosomes in this disease range (37–62 CAG). However, there are no data to validate this mutation rate formula for chromosomes in the intermediate allele range (27–35 CAG), let alone the normal CAG range (<27 CAG). As far as we are aware, there are no data to suggest that all chromosomes with 22 CAG repeats will mutate more rapidly than those with 17 CAG repeats or that the same upward bias exists for normal CAG sizes as for expanded CAG sizes. At high CAG tract sizes, the CAG dependency of the mutation rate may be overwhelming. However, we argue that it is likely that the influence of CAG size on the mutation rate is relatively small at normal-sized CAG tracts, and there is opportunity for other factors to influence stability. The CAG dependency and upward bias of the mutation rate may have a lower cutoff threshold dictated by mechanistic factors such as Okazaki fragment length, DNA damage susceptibility, repair excision tract size, and *cis*-elements.<sup>3,4</sup> Additional experiments are needed to determine whether this large extrapolation of the data from unstable expanded CAG tracts to stable normal-sized CAG tracts is appropriate.

The second assumption of the computational model is that negative selection acts strongly against chromosomes with 50 CAG repeats. In fact, individuals with 50 CAG repeats typically do not become symptomatic until their 30s<sup>5</sup> and would therefore have an opportunity to produce offspring before the onset of disease symptoms. Furthermore, other investigators have observed that adult HD individuals have increased numbers of offspring relative to the general population, and this increased fertility results in positive selection for HD chromosomes.<sup>6</sup> The model also makes assumptions about, but does not incorporate data from, maternally inherited alleles, which are more likely to undergo contractions<sup>7,8</sup> than paternally inherited alleles and would increase the number of generations needed before anticipation results in negative selection. The influence of selection on HD chromosomes is likely not straightforward, and we argue that the computational model overestimates the speed and strength of negative selection on the HD population in the 50–60 CAG range.

The timescale of changes in the computational model is extremely long and may even be longer as a result of overestimating the influence of negative selection. The

computational founder model, which starts with 1/3 of the population (rather than a single founder chromosome) at the 23 CAG size and eliminates chromosomes at 50 CAG, already requires 2,000 generations to form a flattened CAG population distribution (Figure 1 in Falush, bottom left panel, red line) that resembles the data that we observed for variants A1 or A2 (Figure 4C in Warby et al.<sup>1</sup>). Further extending this long timescale (2,000 generations  $\times$  25 years per generation = 50,000 years) may require these founder expansions to occur prior to the colonization of Europe. We agree that it is possible that genetic drift and CAG dependency alone account for the observed pattern in the data, but given the assumptions and the extended amount of time required for this model, it is not the most likely hypothesis.

Instead, we propose that specific haplotypes in the general population have *cis*-element polymorphisms that make CAG expansion more likely compared to other haplotypes. There are multiple stepwise CAG expansions that originate from the general population in these predisposed haplotypes, explaining the unusual similarity between normal and expanded haplotypes as well as the specific enrichment of a few haplotypes on disease chromosomes. The punctuated pattern of disease-associated SNP linkage is due to events that generate new diversity on normal CAG-sized chromosomes containing the predisposing *cis*-element polymorphisms that later undergo CAG expansion. To support this idea, we would agree that the CAG distribution of haplogroup A in the general population of Europe (Figure 3C in Warby et al.,<sup>1</sup> red bars) looks similar to the *cis*-elements model (Figure 1 in Falush, right panel, green haplotype). Furthermore, there is evidence to suggest that specific haplotypes on chromosome 4 (which includes the *HTT* locus) modify the HD age of onset,<sup>9,10</sup> which is inversely related to the CAG tract size and influenced by age-dependent somatic CAG instability.<sup>11</sup>

Finally, it is clear that numerous *cis*-elements influence trinucleotide stability at different loci in the genome. Many genes with similar-sized CAG tracts are found throughout the genome but have different propensities for CAG instability.<sup>12</sup> A *cis*-element adjacent to the *ATXN7* repeat that regulates against CAG hyperinstability has been described.<sup>13</sup> We do not agree with Falush that invoking *cis*-elements that alter particular mutational properties seems unparsimonious, because there is a clear biological precedent for this.<sup>14</sup> *cis*-elements can influence the bias for expansion or contraction<sup>15</sup> and the tissue and developmental specificity of trinucleotide instability.<sup>16</sup> Additionally, *cis*-elements such as CAG tract interruptions can significantly influence the CAG-size dependency of the mutation rate.<sup>17</sup>

With numerous polymorphic satellites in and around the *HTT* gene, as well as SNPs occurring at a frequency of one per 300 bp, we think it is probable that *cis*-element polymorphisms differentially impact CAG stability in the human HD population. Ultimately, further studies are

clearly needed to determine how founder chromosomes, *cis*-elements, *trans*-genetic factors, environmental influences, and stochastic factors contribute to CAG tract instability at the *HTT* locus.

Simon C. Warby,<sup>1</sup> Henk Visscher,<sup>1</sup> Stefanie Butland,<sup>1</sup> Christopher E. Pearson,<sup>2</sup> and Michael R. Hayden<sup>1,\*</sup>

<sup>1</sup>Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, BC V5Z 4H4, Canada; <sup>2</sup>Genetics & Genome Biology Program, The Hospital for Sick Children, Toronto, ON M5G 1L7, Canada

\*Correspondence: [mrh@cmmt.ubc.ca](mailto:mrh@cmmt.ubc.ca)

## Acknowledgments

Funding for this research was provided by the Canadian Genetic Diseases Network, the Huntington Society of Canada, the Michael Smith Foundation for Health Research, the Canadian Institutes of Health Research, the Child & Family Research Institute, the Huntington's Disease Society of America/High Q Foundation, the Muscular Dystrophy Association USA, and the University of Rochester Wellstone Muscular Dystrophy Cooperative Research Center with support from the National Institutes of Health. The authors thank the HD families for the generous use of their DNA for these studies.

## References

1. Warby, S.C., Montpetit, A., Hayden, A.R., Carroll, J.B., Butland, S.L., Visscher, H., Collins, J.A., Semaka, A., Hudson, T.J., and Hayden, M.R. (2009). CAG expansion in the Huntington disease gene is associated with a specific and targetable predisposing haplogroup. *Am. J. Hum. Genet.* **84**, 351–366.
2. Leeflang, E.P., Tavaré, S., Marjoram, P., Neal, C.O., Srinidhi, J., MacFarlane, H., MacDonald, M.E., Gusella, J.F., de Young, M., Wexler, N.S., and Arnheim, N. (1999). Analysis of germline mutation spectra at the Huntington's disease locus supports a mitotic mutation mechanism. *Hum. Mol. Genet.* **8**, 173–183.
3. Pollard, L.M., Sharma, R., Gómez, M., Shah, S., Delatycki, M.B., Pianese, L., Monticelli, A., Keats, B.J., and Bidichandani, S.I. (2004). Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. *Nucleic Acids Res.* **32**, 5962–5971.
4. Cleary, J.D., Nichol, K., Wang, Y.H., and Pearson, C.E. (2002). Evidence of *cis*-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nat. Genet.* **31**, 37–46.
5. Langbehn, D.R., Hayden, M.R., Paulsen, J.S.; and the PREDICT-HD Investigators of the Huntington Study Group. (2009). CAG-repeat length and the age of onset in Huntington disease (HD): A review and validation study of statistical approaches. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* Published online June 22, 2009. 10.1002/ajmg.b.30992.
6. Eskenazi, B.R., Wilson-Rich, N.S., and Starks, P.T. (2007). A Darwinian approach to Huntington's disease: Subtle health benefits of a neurological disorder. *Med. Hypotheses* **69**, 1183–1189.
7. Semaka, A., Collins, J.A., and Hayden, M.R. (2009). Unstable familial transmissions of Huntington disease alleles with 27–35 CAG repeats (intermediate alleles). *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* Published online May 19, 2009. 10.1002/ajmg.b.30970.

8. Wheeler, V.C., Persichetti, F., McNeil, S.M., Mysore, J.S., Mysore, S.S., MacDonald, M.E., Myers, R.H., Gusella, J.F., Wexler, N.S.; , and US-Venezuela Collaborative Research Group. (2007). Factors associated with HD CAG repeat instability in Huntington disease. *J. Med. Genet.* *44*, 695–701.
9. Djoussé, L., Knowlton, B., Hayden, M.R., Almqvist, E.W., Brinkman, R.R., Ross, C.A., Margolis, R.L., Rosenblatt, A., Durr, A., Dode, C., et al. (2004). Evidence for a modifier of onset age in Huntington disease linked to the HD gene in 4p16. *Neurogenetics* *5*, 109–114.
10. Nørremølle, A., Budtz-Jørgensen, E., Fenger, K., Nielsen, J.E., Sørensen, S.A., and Hasholt, L. (2009). 4p16.3 haplotype modifying age at onset of Huntington disease. *Clin. Genet.* *75*, 244–250.
11. Swami, M., Hendricks, A.E., Gillis, T., Massood, T., Mysore, J., Myers, R.H., and Wheeler, V.C. (2009). Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum. Mol. Genet.* *18*, 3039–3047.
12. Butland, S.L., Devon, R.S., Huang, Y., Mead, C.L., Meynert, A.M., Neal, S.J., Lee, S.S., Wilkinson, A., Yang, G.S., Yuen, M.M., et al. (2007). CAG-encoded polyglutamine length polymorphism in the human genome. *BMC Genomics* *8*, 126.
13. Libby, R.T., Hagerman, K.A., Pineda, V.V., Lau, R., Cho, D.H., Baccam, S.L., Axford, M.M., Cleary, J.D., Moore, J.M., Sopher, B.L., et al. (2008). CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: A novel basis for mutational hot spot determination. *PLoS Genet.* *4*, e1000257.
14. Pearson, C.E., Nichol Edamura, K., and Cleary, J.D. (2005). Repeat instability: Mechanisms of dynamic mutations. *Nat. Rev. Genet.* *6*, 729–742.
15. Martins, S., Coutinho, P., Silveira, I., Giunti, P., Jardim, L.B., Calafell, F., Sequeiros, J., and Amorim, A. (2008). Cis-acting factors promoting the CAG intergenerational instability in Machado-Joseph disease. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* *147B*, 439–446.
16. Cleary, J.D., and Pearson, C.E. (2003). The contribution of cis-elements to disease-associated repeat instability: Clinical and experimental evidence. *Cytogenet. Genome Res.* *100*, 25–55.
17. Sobczak, K., and Krzyzosiak, W.J. (2004). Patterns of CAG repeat interruptions in SCA1 and SCA2 genes in relation to repeat instability. *Hum. Mutat.* *24*, 236–247.

DOI 10.1016/j.ajhg.2009.11.006. ©2009 by The American Society of Human Genetics. All rights reserved.