DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases
Conceição Bettencourt, Davina J. Hensman Moss, Michael Flower, Sarah Wiethoff, Alexis Brice, Cyril Goizet, Giovanni Stevanin, Georgios Koutsis, Georgia Karadima, Marios Panas, et al.

To cite this version:
Conceição Bettencourt, Davina J. Hensman Moss, Michael Flower, Sarah Wiethoff, Alexis Brice, et al.. DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. Annals of Indian Academy of Neurology, Medknow Publications, 2016, 10.1002/ana.24656 . hal-01313687

HAL Id: hal-01313687
https://hal.sorbonne-universite.fr/hal-01313687
Submitted on 10 May 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
DNA Repair Pathways Underlie a Common Genetic Mechanism Modulating Onset in Polyglutamine Diseases

Conceição Bettencourt, PhD,1,2 Davina Hensman-Moss, MD,3 Michael Flower, MD,3 Sarah Wiethoff, MD,1,4 Alexis Brice, MD,5,6 Cyril Goizet, MD,7,8 Giovanni Stevanin, PhD,5,9 Georgios Koutsis, MD,10 Georgia Karadima, MD,10 Marios Panas, MD,10 Petra Yescas-Gómez, MD,11 Lizbeth Esmeralda García-Velázquez, MSc,11 María Elisa Alonso-Vilatela, MD,11 Manuela Lima, PhD,12,13,14 Mafalda Raposo, BSc,12,13,14 Bryan Traynor, MD,15 Mary Sweeney, BSc,16 Nicholas Wood, MD,1 Paola Giunti, MD,1,17 Henry Houlden, MD,1,16 Sarah J. Tabrizi, MD,3 and Lesley Jones, PhD18

Objective: The polyglutamine diseases, including Huntington’s disease (HD) and multiple spinocerebellar ataxias (SCAs), are among the commonest hereditary neurodegenerative diseases. They are caused by expanded CAG tracts, encoding glutamine, in different genes. Longer CAG repeat tracts are associated with earlier ages at onset, but this does not account for all of the difference, and the existence of additional genetic modifying factors has been suggested in these diseases. A recent genome-wide association study (GWAS) in HD found association between age at onset and genetic variants in DNA repair pathways, and we therefore tested whether the modifying effects of variants in DNA repair genes have wider effects in the polyglutamine diseases.

Methods: We assembled an independent cohort of 1,462 subjects with HD and polyglutamine SCAs, and genotyped single-nucleotide polymorphisms (SNPs) selected from the most significant hits in the HD study.

Results: In the analysis of DNA repair genes as a group, we found the most significant association with age at onset when grouping all polyglutamine diseases (HD + SCAs; p = 1.43 × 10⁻⁵). In individual SNP analysis, we found significant associations with age at onset when grouping all polyglutamine diseases (HD + SCAs; p = 1.43 × 10⁻⁵).
Over 30 human diseases are caused by expansion of unstable microsatellite sequences. Nine contain repeats that encode glutamine, usually referred to as the polyglutamine diseases (Table 1), and have common features, including autosomal-dominant inheritance (except X-linked spinal and bulbar muscular atrophy), genetic anticipation, neuronal involvement, and intracellular inclusions containing the cognate polyglutamine protein. The phenotypes vary, potentially reflecting differences in the temporal and regional expression and protein context of the disease-causing expansions (see Table 1). There are currently no disease-modifying treatments for these devastating conditions.

In the polyglutamine diseases, longer CAG repeat tracts lead to earlier age at onset (AAO), though the relationship varies between diseases (see Table 1). Not all of the difference in AAO is accounted for by CAG repeat length, and in Huntington’s disease (HD) and at least spinocerebellar ataxia (SCA) types 2 and 3, a substantial portion of this residual variance is heritable, suggesting the existence of additional modifying factors within the genome. The Genetic Modifiers of Huntington’s Disease (GeM-HD) genome-wide association study (GWAS) found two genome-wide loci associated with age at motor onset in HD on chromosomes 15 and 8, with two independent signals at the same locus on chromosome 15. A few SCA genetic modifiers have been proposed and no GWAS have been reported.

Genetic anticipation in these diseases occurs because the repeats are meiotically unstable and tend to expand over successive generations; most also show tissue-specific somatic instability (see Table 1). In HD, somatic instability is expansion-biased and age-dependent, with larger tracts more susceptible to expansion. It occurs in postmitotic neurons and is prominent in striatum and cortex, tissues particularly affected in HD. Somatic instability has been linked to disease onset and progression in both human and mouse HD studies, and decreasing somatic expansion in HD model mice delays phenotype progression. Many of the principles of somatic instability in HD extend to SCAs. Somatic instability has been attributed to the actions of DNA repair proteins, and as well as the individually associated variants, the GeM-HD GWAS found significant association between age at motor onset and several DNA repair pathways. These GeM-HD GWAS findings, along with evidence for somatic instability in other polyglutamine diseases (see Table 1), led us to hypothesize that variants in DNA repair genes might modify AAO in all polyglutamine diseases.

In this report, we demonstrate significant associations between variants in genes involved in DNA repair pathways and the AAO of polyglutamine diseases as a group as well as with some polyglutamine diseases individually.

**Patients and Methods**

**Patients**

Subject cohorts were gathered from the Neurogenetics Unit and Ataxia Center of the National Hospital for Neurology and Neurosurgery (London, UK), TRACK-HD (Europe), SPATAX network (France), the University of Athens Medical School/Eginition Hospital (Athens, Greece), the National Institute of Neurology and Neurosurgery, Manuel Velasco Suarez (Mexico), and the University of Azores (Ponta Delgada, Portugal; Table 2). All subjects with polyglutamine diseases seen at any of the collaborators sites and willing to participate in research were enrolled regardless of their CAG repeat size or AAO. All studies were approved by local ethics committees, and all subjects gave written informed consent. For this study, we gathered samples and data for HD and SCAs 1, 2, 3, 6, 7, and 17; very few dentatorubral-pallidoluysian atrophy (DRPLA) and spinal and bulbar muscular atrophy (SBMA) samples were available to us, so these diseases were not included. AAO and CAG repeat size was available for 1,462 patients (see Table 2). Given the varied phenotypes of polyglutamine diseases, motor onset (HD) or onset of the first progressive symptom as reported by the patient was used to determine AAO throughout all cohorts. Given the small number of patients, SCA17 was only considered in the combined SCA analysis.
TABLE 1. Characteristics of the Polyglutamine Diseases

<table>
<thead>
<tr>
<th>Repeat Disorder</th>
<th>Gene</th>
<th>Prevalence</th>
<th>Variance in AAO Explained by Repeat Length</th>
<th>Normal Range</th>
<th>Pathogenic Range</th>
<th>Somatic Instability</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD</td>
<td>HTT</td>
<td>3–10</td>
<td>50–60% (40–60%)</td>
<td>6–35</td>
<td>40–121</td>
<td>Yes</td>
</tr>
<tr>
<td>SCA1</td>
<td>ATXN1</td>
<td>0.16</td>
<td>64–76% (no detected heritable component)</td>
<td>6–38</td>
<td>45–83</td>
<td>Yes</td>
</tr>
<tr>
<td>SCA2</td>
<td>ATXN2</td>
<td>0.2</td>
<td>50–80% (17–59%)</td>
<td>15–31</td>
<td>33–500</td>
<td>Yes</td>
</tr>
<tr>
<td>SCA3</td>
<td>ATXN3</td>
<td>0.4</td>
<td>45–80% (46%)</td>
<td>12–44</td>
<td>52–87</td>
<td>Yes</td>
</tr>
<tr>
<td>SCA6</td>
<td>CACNA1A</td>
<td>0.04</td>
<td>26–52% (no detected heritable component)</td>
<td>4–18</td>
<td>20–33</td>
<td>Unknown</td>
</tr>
<tr>
<td>SCA7</td>
<td>ATXN7</td>
<td>0.12</td>
<td>71–88% (no detected heritable component)</td>
<td>3–19</td>
<td>37–460</td>
<td>Yes</td>
</tr>
<tr>
<td>SCA17</td>
<td>TBP</td>
<td>&lt;0.02</td>
<td>Unknown</td>
<td>25–40</td>
<td>49–66</td>
<td>Unknown</td>
</tr>
<tr>
<td>DRPLA</td>
<td>ATN1</td>
<td>0.005–0.04</td>
<td>50–68%</td>
<td>6–35</td>
<td>48–93</td>
<td>Yes</td>
</tr>
<tr>
<td>SBMA</td>
<td>AR</td>
<td>0.65–2.00</td>
<td>29%</td>
<td>9–34</td>
<td>38–72</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Epidemiology and CAG repeat ranges of polyglutamine diseases. Prevalence is given/100,000 European population. AAO = age at onset; HD = Huntington’s disease; SCA = spinocerebellar ataxia; DRPLA = dentatorubral-pallidoluysian atrophy; SBMA = spinal and bulbar muscular atrophy.

significant gene-wide *p* values (see Table S5 of the GeM-HD article). For each gene, the most significant SNP was selected, along with a small number of proxy SNPs in close LD (*r*² > 0.8) with the most significant SNP that also showed association in GeM-HD. Where possible, these proxy SNPs were chosen to have functional annotation (http://browser.1000genomes.org/index.html: accessed 12/6/14). If a gene contained two independent significant signals in GeM-HD (e.g., FAN1), then the lead SNP for the second signal was included. Note that this selection procedure is not intended to give comprehensive

TABLE 2. Cohort Characteristics

<table>
<thead>
<tr>
<th>Cohort</th>
<th>HD</th>
<th>SCA1</th>
<th>SCA2</th>
<th>SCA3</th>
<th>SCA6</th>
<th>SCA7</th>
<th>SCA17</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Athens, Greece</td>
<td>351</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>351</td>
</tr>
<tr>
<td>Azores, Portugal</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>91</td>
</tr>
<tr>
<td>London, UK</td>
<td>0</td>
<td>30</td>
<td>66</td>
<td>45</td>
<td>69</td>
<td>7</td>
<td>1</td>
<td>218</td>
</tr>
<tr>
<td>Mexico</td>
<td>0</td>
<td>0</td>
<td>113</td>
<td>0</td>
<td>0</td>
<td>66</td>
<td>6</td>
<td>185</td>
</tr>
<tr>
<td>SPATAX, France</td>
<td>0</td>
<td>147</td>
<td>115</td>
<td>261</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>523</td>
</tr>
<tr>
<td>TRACK-HD, Europe</td>
<td>94</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>94</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>445</td>
<td>177</td>
<td>294</td>
<td>397</td>
<td>69</td>
<td>73</td>
<td>7</td>
<td>1,462</td>
</tr>
</tbody>
</table>

% M = percentage of males; AAO = age at onset; SD = standard deviation.

*One subject had no sex information.
HD = Huntington’s disease; SCA = spinocerebellar ataxia.
coverage of the genes in question, but instead to highlight SNPs likely to be disease relevant. To guard against the effects of population stratification, SNPs were removed from the analysis if they had a Hardy-Weinberg p value <0.001 in the whole data set. These procedures yielded 22 genotyped SNPs with success rates ranging from 94.2% to 98%, as described in Supplementary Table 1.

SNP genotyping was performed using custom KASP assays at LGC Genomics (Hertfordshire, UK). Gene-level sense sequences were used to design SNP assays (see Supplementary Table 2). The assays for several SNPs were designed in reverse orientation to the chromosome (rs4150407, rs1805323, rs1037700, rs1037699, rs3512, and rs20579). For this reason, for all SNPs in reverse orientation to the chromosome (rs4150407, rs1805323, rs1037700, rs1037699, rs3512, and rs20579), genotypes resulting from these KASP assays will be complementary to those using HGVS nomenclature. This is reflected in Supplementary Table 3, where the minor allele for these SNPs differs from GeM-HD, but corresponds to the same allele.

### Statistical Analyses

AAOs for the various diseases were corrected for repeat length using a similar method to the GeM-HD GWAS. A linear regression was performed for each disease separately of ln(AAO) on expanded repeat length. Regression parameters are given in Table 3. These parameters were used to construct an expected value of AAO for each individual, based on their repeat length, which was subtracted from their actual AAO to give a residual. Association of each SNP with AAO was tested by performing a regression of these residuals on the number of minor alleles. As a secondary analysis, individual SNP associations were examined. Three of these were significant after Bonferroni correction for eight tests) were observed for HD+SCAs (p = 1.43 × 10−5), HD (p = 0.00194), all SCAs (p = 0.00107), SCA2 (p = 0.00350), and SCA6 (p = 0.00162). The increased significance of these associations compared to an undirected test using two-sided SNP p values (see supplementary Table 3) indicates concordance in the direction of effects across SNPs between these samples and GeM-HD. In particular, the observed association with HD is a convincing replication of the GeM-HD results in an independent sample.

### Results

In the primary analysis, which tested the overall effect of all 22 SNPs on AAO, significant associations (after Bonferroni correction for eight tests) were observed for HD+SCAs (p = 1.43 × 10−5), HD (p = 0.00194), all SCAs (p = 0.00107), SCA2 (p = 0.00350), and SCA6 (p = 0.00162). The increased significance of these associations compared to an undirected test using two-sided SNP p values (see supplementary Table 3) indicates concordance in the direction of effects across SNPs between these samples and GeM-HD. In particular, the observed association with HD is a convincing replication of the GeM-HD results in an independent sample.

As a secondary analysis, individual SNP associations were examined. Three of these were significant after Bonferroni correction for eight disease combinations and 22 SNPs (Table 4 and Supplementary Table 4): rs3512 in FAN1 with all SCAs and HD+SCAs and rs1805323 in PMS2 with HD+SCAs. Each association was in the same direction as in GeM-HD. We did not replicate the most significant signal in GeM-HD, rs146353869 (p = 4.30 × 10−20, associated with 6 years earlier age at motor onset of HD). This is likely due to our sample being much smaller than GeM-HD and thus less well powered to find associations with SNPs with relatively low-frequency minor allele frequency (MAF) such as rs146353869 (MAF = 0.017). However, rs3512, the most significant individual SNP in this study, indexes the second significant chromosome 15 signal in GeM-HD (p = 5.28 × 10−13, associated with 1.4 years later onset of HD), and is in the 3’UTR (untranslated region) of FAN1. Three SNPs (rs1037700, rs5893603, and rs16869352) were found to be in high LD (r² > 0.8) in our sample with more significant SNPs from GeM-HD. Removing these SNPs reduced the significance of the multi-SNP associations with SCA2 and SCA6, although...
these remained nominally significant (see Supplementary Table 3). Finally (see Supplementary Table 3), all the significant multi-SNP associations from the primary analysis remained significant after removing the most significant single SNP (rs3512), suggesting that the signal enrichment is not being driven by a single SNP.

To visualize the combined effect of our SNPs on residual AAO a polygenic "age at onset score" was derived, defined as the sum of the number of minor alleles at each locus weighted by their effect size in GeM-HD (note that negative scores here correspond to earlier AAO). The residual AAO for each quartile of this risk score was plotted in Figure 2. As expected, there was a positive correlation between residual AAO in our data and increasing age at onset score, although the effect was small—the score accounts for approximately 1% of the variance of residual AAO.

**Discussion**

Our data implicate a common mechanism by which genetic variation in DNA repair pathways underlies age at onset of disease in multiple polyglutamine diseases. Alterations in DNA repair pathways could predispose to earlier onset by interacting with polyglutamine etiology at various levels. Rare loss-of-function variants in DNA repair genes cause multiple recessive ataxias; \(^{22}\) \(ATM\) encodes a master regulator of DNA repair following double-strand breaks; \(^{23}\) \(PNPK\) encodes a DNA-specific kinase that facilitates DNA repair; \(^{24}\) \(APTX\) encodes a protein that interacts with PARP1 to mediate single strand DNA breaks; \(^{25}\) and mutations in \(TDP1\) also give defects in single-strand break repair. \(^{26}\) The mechanisms by which neurodegeneration and ataxia result from these losses of function are not conclusively established, but there is substantial evidence for the fine control exercised by \(ATM\) being critical in cell division and cell death pathways, which could lead to neuronal loss. \(^{27}\) However, it is notable that none of the genes associated with recessive ataxia syndromes were identified to contain HD-related variants in HD-GeM.\(^6\)

Repetitive DNA sequences can form unusual secondary structures \(^{28}\) to which DNA mismatch repair proteins bind, and, in the process of repair, cause somatic instability (often expansion) of the CAG repeats. A number of enzymes with the ability to nick DNA and therefore necessitate DNA repair are known to promote CAG expansion and both somatic expansion and HD-related phenotypes are ameliorated in mouse models by manipulating genes associated with DNA repair. \(^{15,29–32}\) Critically, delay in phenotype onset in HD mice was recently demonstrated through suppressing somatic expansion by crossing HTT knock in mice with \(Ogg1–/–\) mice, lacking the DNA cleaving 7,8-dihydro-8-oxo-guanine (8-oxo-G) glycosylase. \(^{16}\) Notably, the single most significant SNP in the present study, rs3512, is in the 3'UTR of \(FAN1\), which has DNA endo/exonuclease activity. Larger CAG repeats are associated with more-severe pathology and earlier disease onset in affected patients; therefore, somatic expansion provides a plausible mechanism by which the genetic variation we identify here could alter AAO of polyglutamine diseases (Fig 1). Additional consequences of impaired DNA repair cannot be discarded though (Fig 1A), and these may also be implicated in a wider range of neurodegenerative diseases, including several ataxic syndromes. \(^{33,34}\)

### Table 4. Individual SNPs Showing Significant Associations With the Age at Onset in Polyglutamine Diseases

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Disease Group</th>
<th>Two-Sided (p) Value</th>
<th>Same Direction as GeM-HD?</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3512</td>
<td>FAN1</td>
<td>All (HD+SCAs)</td>
<td>(1.52 \times 10^{-5})</td>
<td>Yes</td>
</tr>
<tr>
<td>rs1805323</td>
<td>PMS2</td>
<td>All (HD+SCAs)</td>
<td>(3.14 \times 10^{-5})</td>
<td>Yes</td>
</tr>
<tr>
<td>rs3512</td>
<td>FAN1</td>
<td>All SCAs</td>
<td>(2.22 \times 10^{-4})</td>
<td>Yes</td>
</tr>
</tbody>
</table>
| **Significant After Bonferroni for 22 SNPs and 8 Disease Groupings \((p < 2.84 \times 10^{-4})**
| rs1805323 | PMS2 | HD                   | \(3.14 \times 10^{-4}\) | Yes                        |
| rs1805323 | PMS2 | SCA1                 | \(1.67 \times 10^{-3}\) | Yes                        |
| rs1037699 | RRM2B| SCA6                 | \(4.86 \times 10^{-4}\) | Yes                        |
| rs1037700 | RRM2B| SCA6                 | \(5.47 \times 10^{-4}\) | Yes                        |
| rs5893603 | RRM2B| SCA6                 | \(2.13 \times 10^{-3}\) | Yes                        |

HD = Huntington’s disease; SCA = spinocerebellar ataxia.
There are several issues likely to have reduced the power of our study. The sample sizes for many of the SCA were relatively small, and despite modeling the relationship of age of onset to CAG length separately for each disease, there is likely to be heterogeneity between diseases in this and potentially other respects that we have not been able to consider. We could not account for interruptions of pure CAG repeat tracts, which may stabilize repeat instability; thus, our power to detect any effects mediated by somatic instability may have been reduced. Nevertheless, we have shown that DNA repair genes as a group significantly modify AAO in the polyglutamine diseases taken together, in HD, in all SCA, SCA2, and SCA6. Additionally, we have identified potential modifier SNPs in HD, SCA1, and SCA6 (Table 4 and Supplementary Table 4). The effects of these SNPs on AAO are quite small, and it would be worth repeating the analysis with larger samples and more SNPs as the predictive power of such polygenic risk scores increases as sample size and number of variants genotyped increase.

By suggesting common mechanisms for polyglutamine diseases, our findings offer novel therapeutic opportunities in multiple diseases along with the potential to improve clinical trial design by stratifying subject variability. Molecules targeting DNA repair have been...
developed and are used in the clinic to treat cancers,\textsuperscript{39,40} and such therapeutics, along with others in development, may prove useful in some or all of the polyglutamine diseases. Furthermore, these shared mechanisms may extend to diseases associated with non-CAG and nontranslated repeats, most likely in those that show somatic instability.

Acknowledgment

This work was undertaken at UCLH/UCL who acknowledges support from the Department of Health’s NIHR Biomedical Research Centre. This work was supported by the European Commission 7th Framework Program (FP7/2007-2013) under grant agreement no. 2012-305121 “Integrated European –omics research project for diagnosis and therapy in rare neuromuscular and neurodegenerative diseases (NeurOmics)”, the Medical Research Council (MRC UK, MR/L010305/1), the Wellcome Trust in equipment and strategic award (Synaptopathies) funding (WT093205MA and WT104033/Z/14/Z), the Brain Research Trust (BRT), the MSA Trust, and the National Institute for Health Research (NIHR) University College London Hospitals (UCLH) Biomedical Research Centre.

We are grateful to all those who contributed samples and patients for their kind cooperation and to the patients, their families, and carers. We thank M.H. Parkinson for collection of samples at the Ataxia Center. Many thanks to the DNA and cell bank of the ICM in Paris for sample preparation. We thank Michael O’Donovan for critical reading of the manuscript.

Author Contributions


Potential Conflicts of Interest

The author declare no conflicts of interest.

References


