GENETIC MODIFERS OF HUNTINGTON’S DISEASE

ELIANA MARISA DA SILVA RAMOS
TESE DE DOUTORAMENTO APRESENTADA
AO INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
DA UNIVERSIDADE DO PORTO EM
CIÊNCIAS BIOMÉDICAS
GENETIC MODIFIERS OF HUNTINGTON’S DISEASE

Tese de Candidatura ao grau de Doutor em Ciências Biomédicas submetida ao Instituto de Ciências Biomédicas de Abel Salazar, Universidade do Porto

Orientador – Prof. Doutora Isabel da Conceição Moreira Pereira Alonso
Professora do Instituto de Ciências Biomédicas Abel Salazar

Co-orientador – Prof. Doutora Marcy Elizabeth MacDonald
Professora da Harvard Medical School

Co-orientador – Prof. Doutor António Jorge dos Santos Pereira de Sequeiros
Professor Catedrático do Instituto de Ciências Biomédicas Abel Salazar

Porto
2013
“But even if we take it for granted that we have discovered the sedes morbi we are still left in ignorance in regard to the nature of the derangement. And here we must leave the interesting subject of the pathology of this disease, and trust that the science, which has accomplished such wonders, through the never-tiring devotion of its votaries, may yet “overturn and overturn, and overturn it,” until it is laid open to the light of day.”

George Huntington, M.D., 1872
# Table of Contents

Publications ............................................................................................................. i
Abbreviations .......................................................................................................... iii
Abstract .................................................................................................................. v
Sumário ..................................................................................................................... vii
Introduction ............................................................................................................. 1

## Clinical Presentation ....................................................................................... 3
- Motor symptoms .................................................................................................. 3
- Psychiatric symptoms ......................................................................................... 4
- Cognitive symptoms ........................................................................................... 4
- The atypical phenotypes ................................................................................. 4
- Neuropathology ................................................................................................. 5

## The genetic cause .......................................................................................... 6
- CAG repeat distribution ..................................................................................... 6
- CAG repeat length and age-at-onset ................................................................. 7
- CAG repeat instability ....................................................................................... 9

## The origin of Huntington’s disease ............................................................... 9
- Prevalence of Huntington’s disease ................................................................ 10
- Haplotypes studies .......................................................................................... 11

## Genetic modifiers ......................................................................................... 12
- Candidate gene association studies ................................................................. 13
  - Cis-modifiers .................................................................................................. 13
  - Trans-modifiers ............................................................................................. 14
- Unbiased scans for genetic modifiers ............................................................... 15

## The Huntington’s disease model systems .................................................. 18
- The mouse model ............................................................................................. 18
  - Hdh^{Q111} knock-in mouse model ................................................................ 19

## Major Goals .................................................................................................... 21
## Specific objectives .......................................................................................... 23
Publications

In this thesis were used results, already published or in preparation, as described below:


Abbreviations

AD – Alzheimer’s disease
ALS – amyotrophic lateral sclerosis
AO – age-at-onset
ATN1 – atrophin1
ATXN1 – ataxin1
ATXN2 – ataxin2
ATXN3 – ataxin3
BP – Bipolar disease
CACNA1A - calcium channel, voltage-dependent, P/Q type, alpha 1A subunit
Chr – chromosome
COHORT - Cooperative Huntington Observational Research Trial
COMT – catechol-O-methyltransferase
CSS – chromosome substitution strain
DARPP-32 - Dopamine- and cAMP-regulated phosphoprotein, Mr 32 kDa
DAT1 – dopamine active transporter 1
DNA – deoxyribonucleic acid
DRD2 – dopamine receptor D2
DRD4 – dopamine receptor D4
DRPLA - Dentatorubral pallidoluysian atrophy
EHDN – European Huntington’s Disease Network
GABA – gamma-aminobutyric acid
GEE – generalized estimated equation
GRIN2A – Glutamate Receptor Ionotropic N-Methyl D-Aspartate 2A
GRIN2B – glutamate Receptor Ionotropic N-Methyl D-Aspartate 2B
GRIK2 – glutamate receptor ionotropic kainate 2
GWA – genome-wide association
HD – Huntington’s disease
HD-MAPS – Huntington’s Disease Modifiers in Age of Onset in Pairs of Siblings
HTT – huntingtin
IQR – interquartile
MAF – minor allele frequency
MDD – major depressive disorder
MSN – medium spiny neurons
n – number
NA – not ascertained
NII – neuronal intranuclear inclusion
NMDA – N-methyl-D-aspartate
PCA – principal component analysis
PCR – polymerase chain reaction
PD – Parkinson’s disease
PolyQ – polyglutamine
PPARGC1A – peroxisome proliferator receptor gamma coactivator 1 alpha
QTL – quantitative trait loci
SCA – spinocerebellar ataxia
SD – standard deviation
SNP – single nucleotide polymorphism
STR – single tandem repeat
TBP – TATA box binding protein
UTR - untranslated region
Abstract

Huntington’s disease (HD) is an autosomal dominant neurodegenerative disorder characterized by motor, cognitive and behavioral disturbances, in which all affected individuals have precisely the same type of mutation, an expansion of a polymorphic CAG repeat in the coding region of the \textit{HTT} gene. The mutant gene is present since conception initiating a lifelong pathogenic process that leads to onset of symptoms typically in mid-life. Even though there is a strong inverse correlation between onset of HD clinical symptoms and CAG repeat size, patients with precisely the same length of the expanded CAG repeat may differ dramatically in age-at-onset (AO). Several studies have shown that this deviation in motor AO is highly heritable, strongly implying the existence of stochastic, environmental and/or genetic factors. Understanding the mechanisms modulating HD expressivity is a major challenge in the HD field, whose ultimate goal is the development of effective therapeutic strategies. Therefore the aim of this doctoral project was to assess, through several classical genetic approaches, the relative contribution of genetic factors distinct from the CAG mutation that modify the disease outcome.

Detailed haplotype analyses of two distinct European sub-populations has refined our understanding of the origins of the \textit{HTT} CAG mutation, but more importantly, has demonstrated that 4p16.3 haplotypes have no influence on the HD pathogenesis or the \textit{HTT} repeat instability. Moreover, our rigorous analysis of over 3,500 HD subjects showed that AO of motor manifestations is not determined by the normal allele or the interaction between expanded and normal alleles, corroborating our haplotype findings. Additionally, our critical studies of candidate gene modifiers of residual AO have demonstrated that when proper attention is paid to ancestry and, importantly, proper estimation of residual phenotype, there is no evidence for a significant effect of any of the consistently reported common variants at the most compelling candidates (\textit{GRIK2, PPARGC1A, GRIN2A, GRIN2B}) or of the functional variants in genes of the dopaminergic pathway (\textit{DAT1, DRD2, DRD4, COMT}).

There is currently suggestive evidence of CAG repeat polymorphisms contributing to complex traits and even acting as modifiers of other diseases, including previous work from our group that has shown that reduced penetrance CAG alleles are over-represented in diagnosed cases of major depressive disorder (MDD). Our data showed that the functional \textit{HTT} CAG repeat does not act as a sensitizer to the effects of the risk factors in
the general population that contribute to the manifestation of amyotrophic lateral sclerosis (ALS) or bipolar disorder (BP), implying that the CAG repeat contributes to MDD and perhaps other complex disorders, but not ALS or BP.

Finally, we have evaluated a novel human-mouse cross-species approach to test the functional effect of a region, emerging from linkage analysis in HD families, which appears to harbor loci modifying AO of motor symptoms. We specifically asked whether dominantly acting A/J-C57BL6/J variants at this region could influence CAG repeat instability and/or the rate of the early disease process that leads to pathology in heterozygous C57BL6/J Hdh^Q111^ knock-in mice. We found suggestive evidence that this specific loci is functionally modulating the levels and pattern of somatic CAG instability in liver and the appearance of intranuclear inclusions in the striatum, as well as DARPP-32 levels in medium spiny striatal neurons, encouraging a larger study to gather additional information that will allow us to better understand which of the candidate genes in this region is indeed modifying these traits.

In conclusion, this work has provided the basis for a more rigorous and reliable statistical assessment of modifier variants in the HD research community and has set up the stage for the use of a novel functional approach to map modifiers by parsing candidate regions. More importantly, our powerlessness to single out compelling candidate modifier genes of HD reinforces the need of genome-wide screens for common and rare variants to identify strong and convincing genetic modifiers of HD pathogenesis.
Sumário

A doença de Huntington é uma doença neurodegenerativa com transmissão autossômica dominante caracterizada por alterações motoras, cognitivas e comportamentais, em que todos os indivíduos afetados têm exatamente o mesmo tipo de mutação, a expansão de uma repetição CAG na região codificante do gene HTT. O gene mutado está presente desde concepção iniciando um processo patogênico que se mantém ao longo da vida e que resulta no aparecimento dos sintomas normalmente na idade adulta. Apesar de existir uma forte correlação inversa entre o início dos sintomas clínicos da doença de Huntington e o tamanho da repetição de CAG, doentes exatamente com o mesmo tamanho de expansão podem diferir dramaticamente na idade de início. Vários estudos têm demonstrado que este desvio na idade de início dos sintomas motores é altamente hereditário, sugerindo a existência de factores estocásticos, ambientais e/ou genéticos. Compreender os mecanismos que modulam a expressividade da doença de Huntington é um grande desafio nesta área, em que o objetivo final é o desenvolvimento de estratégias terapêuticas eficazes. Este projeto de doutoramento teve como objetivo avaliar, recorrendo a várias metodologias da genética clássica, a contribuição relativa de outros fatores genéticos, para além da mutação, que modificam a progressão da doença de Huntington.

A análise detalhada de haplótipos em duas subpopulações europeias distintas permitiu um conhecimento mais detalhado sobre as origens da mutação no gene HTT, e demonstrou que os haplótipos na região 4p16.3 não têm qualquer influência na patogénese da doença ou na instabilidade da repetição. Além disso, a análise rigorosa de mais de 3500 doentes de Huntington revelou que a idade em que surgem os primeiros sintomas motores não é determinada nem pelo alelo normal nem pela interação entre os alelos normal e expandido, corroborando os resultados obtidos pelo estudo de haplótipos. A análise crítica de genes candidatos modificadores da idade de início residual demonstraram que quando se têm em consideração a ancestralidade das amostras e uma estimativa correta da idade de início residual, não há nenhuma evidência de um efeito significativo em qualquer uma das variantes comuns nos genes candidatos com maior evidência na literatura (GRIK2, PPARGC1A, GRIN2A, GRIN2B) ou em variantes funcionais em genes da via dopaminérgica (DAT1, DRD2, DRD4, COMT).
Existem evidências que sugerem que repetições do tripleto CAG contribuem para traços complexos e podem até agir como modificadores de outras doenças, tal como demonstrado num estudo prévio do nosso grupo em que se verificou que alelos de penetrância reduzida estão sobrerepresentados em indivíduos diagnosticados com transtorno depressivo major. Os nossos resultados demonstraram que a repetição funcional no gene HTT não atua como um sensibilizador para os efeitos dos fatores de risco que contribuem para a manifestação da esclerose lateral amiotrófica ou da doença bipolar na população em geral, mas que a repetição contribui para transtorno depressivo major e talvez para outras doenças complexas, que não para a esclerose lateral amiotrófica ou a doença bipolar.

Finalmente, foi avaliada uma nova abordagem entre espécies humano-ratinho para testar o efeito funcional de uma região, que foi detetada a partir de análise de ligação genética em famílias de doentes com Huntington, que aparentemente inclui loci modificadores da idade de início de sintomas motores. Especificamente, questionou-se se variantes A/J-C57BL6/J dominantes poderiam influenciar a instabilidade da repetição e/ou a progressão dos processos envolvidos nas primeiras etapas da doença e que resultam na patologia observada em ratinhos knock-in C57BL6/J HdhQ111 heterozigotos. Encontramos evidências que sugerem que este loci específico está a modular funcionalmente os níveis e padrões de instabilidade somática no fígado e o aparecimento de inclusões intranucleares no estriato, bem como os níveis de DARPP-32 nos neurônios espinhosos médios, justificando um estudo mais alargado que nos permitirá entender melhor qual dos genes candidatos nesta região está de fato a modificar estas características.

Em conclusão, este trabalho constitui um trabalho seminal para a comunidade de investigação da doença de Huntington contribuindo para uma avaliação estatística mais rigorosa de variantes modificadoras e representa o ponto de partida para a utilização de uma nova abordagem funcional para mapear genes modificadores através da análise das regiões candidatas em ratinhos. Mais importante, a nossa dificuldade em identificar genes modificadores da doença de Huntington reforça a necessidade de estudos genómicos para a pesquisa de variantes raras e comuns que contribuam como modificadores genéticos robustos da patogénesis da doença de Huntington.
Introduction
In 1872, Dr. George Huntington, an American physician, described the distinguished chorea features that he observed in families from Long Island, New York. Even though choreic movement disorders had already been well documented, Dr. Huntington believed that the late onset and hereditary nature of the choreic features he observed in these families set it apart from previously described chorea. This ended up being the first exhaustive description of a specific disorder entity that was originally called “Huntington’s chorea” in large part because of Dr. Huntington’s detailed and accurate definition of the disease in his brief paper [1]: “It is attended generally by all the symptoms of common chorea, only in an aggravated degree, hardly ever manifesting itself until adult or middle life, and then coming on gradually but surely, increasing by degrees, and often occupying years in its development, until the hapless sufferer is but a quivering wreck of his former self.” For many decades its name remained unchanged, until the nineteen-eighties when the name changed to “Huntington’s disease” (HD) to reflect the fact that chorea is not the only manifestation of the disease.

Clinical Presentation

HD is a devastating inherited neurodegenerative disorder with a wide spectrum of signs and symptoms characterized primarily by progressive motor, cognitive and psychiatric disturbances. The clinical diagnosis of HD is mostly based on the occurrence of chorea. Other prevalent and often debilitating features of HD include weight loss [2-6] and sleep disturbance [7-10]. Early signs of the disease vary greatly from person to person; occurring typically in mid-life, between the ages of 30 and 50, though the disorder can manifest at any age. The average lifespan after onset is 15 to 20 years, with the most common cause of death being pneumonia and heart disease [11-14].

Motor symptoms

The first motor signs of HD are subtle: minor clumsiness, difficulties with smooth eye pursuit and slight uncontrolled or awkward movements that progress slowly over a period of years. These movement abnormalities become more and more exaggerated until the entire body is consumed with involuntary jerking or writhing movements, commonly known as chorea. While chorea typically dominates in early phases of the disease other movement abnormalities, including dystonia (an involuntary sustained contracture of
muscles), muscle rigidity and bradykinesia (slowness in the execution of the movement), become more marked in late stages of the disease. Other motor symptoms may include slow uncoordinated fine movements; slow or abnormal eye movements; impaired gait, posture and balance; speech impairment and difficulty in swallowing.

**Psychiatric symptoms**

Psychiatric features are very frequent in the early stages of the disease and often precede motor onset by many years [15]. The most frequent psychiatric sign is depression, with about 40% of HD patients having symptoms of depression [16]. Also, the suicide rate is dramatically increased in those “at risk” for HD compared to the general population [17]; with 17.5% of the carriers of the HD mutation having suicidal thoughts, from whom 10% had made at least one suicide attempt in the past [16]. While obsessive-compulsive thoughts and repetitive behaviors are relatively common in HD, psychotic symptoms are rarer though delusions and hallucinations can occur, mainly in the later stages of the disease [18-20]. A wide range of other neuropsychiatric symptoms, such as dysphoria, agitation, irritability, apathy, anxiety, disinhibition are also characteristic of individuals with HD [19].

**Cognitive symptoms**

The cognitive capacities are also severely affected, and like psychiatric symptoms, deficits in cognitive functions can be detected many years before the onset of motor symptoms [21]. The cognitive decay is mostly related to executive functions, with patients reporting difficulties with multi-tasking and concentration. In HD patients, memory and learning certainly becomes impaired, and these individuals also have psychomotor symptoms and subtle visuospatial perceptual changes [22]. As cognitive dysfunction progresses, patients can develop severely limiting dementia.

**The atypical phenotypes**

While the majority of HD patients have the typical mid-life onset hyperkinetic phenotype, there are two extreme variants of phenotypes: juvenile and late onset HD. In juvenile HD the first symptoms and signs of the disease start before the age of 20 years. It is uncommon, only representing fewer than ten percent of individuals with HD [23], and differs from typical adult onset as the movement disorder is typically hypokinetic rather than hyperkinetic: including primarily tremors, bradykinesia, muscular rigidity and dystonia,
while chorea may be completely absent [24-28]. In juvenile HD, seizures are seen in 30-50% of the cases [29-32]; and cerebellar signs, myoclonus and spasticity may also occur [33,34]. On the other hand, in late onset HD, the first symptoms do not start until the individual is over 60 years old. The signs are predominantly motor disturbance with relatively mild disability and slower progression, with dementia sometimes associated with coincident Alzheimer's disease (AD) [35].

**Neuropathology**

The clinical symptoms of HD are due to the gradual atrophy of the striatum, where neuropathological analyses of patients' brains have revealed a selective and progressive loss of medium spiny neurons (MSNs) in the caudate nucleus and the putamen [36].

The MSNs neurons represent 90-95% of the striatal neuronal population playing a key role in initiating and controlling movements of the body, limbs and eyes. They are innervated by dopaminergic inputs originated from the substantia nigra and glutamatergic inputs from the cortex [37] and project back to the basal ganglia output nuclei via a direct and indirect pathway. These neurons use the inhibitory neurotransmitter gamma-aminobutyric acid (GABA) and while neurons of the direct pathway express substance P and dopamine D1 receptors (excitatory effect), neurons of the indirect pathway express enkephalin and dopamine D2 receptors (inhibitory effect). Enkephalin-containing neurons projecting to the external segment of the globus pallidus are preferentially lost in postmortem brains of symptomatic and presymptomatic individuals while substance P-containing neurons projecting to the internal pallidal segment are relatively spared [38-40]. This imbalance between the direct and indirect pathway and consequent loss of the GABAergic MSNs inhibitory input is believed to be the underlying cause of the uncontrolled movements characteristic of HD [reviewed in 41].

Even though the loss of MSNs is the earliest and most evident pathological sign of HD, neuronal loss also occurs throughout the cerebral cortex, particularly in layers III, V and VI, and as the disease progresses other regions may be affected, with HD individuals overall brain weight decreasing up to 40% by the time of death [reviewed in 42]. Indeed, imaging studies of pre-manifest and early manifest HD patients revealed that substantial cortical and striatal atrophy occurs at very early stages, and the latter can even occur years prior to the expected onset [43-46].

Another characteristic feature of HD is the presence of neuronal intranuclear inclusions (NIIs) and cytoplasmatic inclusions, containing aggregated expanded huntingtin and other proteins, in striatal and cortical neurons [47-50]. The pathogenic relevance of these
huntingtin aggregates is an ongoing debate in the HD field, as it is still unclear whether they cause neuronal death by disrupting the normal transcriptional process in the susceptible neurons [51], whether they are sequestrating the mutant protein to hinder polyglutamine-induced neurotoxicity [52-54] or whether their presence is just neutral to the disease process.

The genetic cause

In 1983, using standard linkage analyses in a large Venezuelan kindred with apparent high HD prevalence and in multigenerational American HD families, the huntingtin (HTT) gene was mapped to the tip of the short arm of chromosome 4 [55]. It took the Huntington’s Disease Collaborative Research Group another 10 years to identify the underlying mutation that causes HD [56], making this the first autosomal disease locus found using genetic linkage analysis. Using haplotype analyses of linkage disequilibrium in 75 HD families of different origins, the precise causal gene at 4p16.3 was isolated and showed the presence of a polymorphic trinucleotide CAG repeat in the coding region of the HTT gene that was expanded on all HD families examined [56]. HD was ultimately found to be caused by this expanded and unstable CAG repeat, which encodes an elongated polyglutamine (polyQ) tract near the amino terminus of the huntingtin protein, a large (~349 kDa) widely expressed protein with unknown function.

Predictive testing of HD started being offered to “at-risk” individuals in 1988, first with use of a linked DNA marker [57], and later on, with the discovery of the genetic mutation, it was possible to offer accurate tests for HD by determining the HTT CAG repeat length in these individuals. Not only “at-risk” individuals now have the opportunity to learn whether or not they have inherited an expanded repeat that will cause HD, but can also make an informed decision about prospective pregnancies through prenatal testing. However, predictive testing involves significant psychological and social challenge and, nowadays there are still approximately 85% of “at-risk” HD individuals that do not come forward for testing [58-61].

CAG repeat distribution

Normal HTT alleles are polymorphic, having fewer than 35 CAG repeats and are considered to produce no overt clinical symptoms of HD. These alleles can be grouped in two categories: normal stable (with less than 27 CAGs) and high-normal alleles (27 to 35 CAGs) (Figure 1). The average CAG repeat size in the general population is 16 to 20
repeats, depending on its geographic origin [62], and the repeat length distribution is positively skewed, with more alleles lying above than below the modal length [63].

Figure 1 – The HTT CAG repeat distribution in HD families. Bimodal distribution of the CAG repeat length, with a clear peak in the normal (17 CAG repeats) and other in the full penetrance (42-43 CAG repeats) range. HD stable normal alleles are represented in blue; high-normal in green; reduced-penetrance in orange and full penetrance in red (lighter red represents the allele sizes normally associated with juvenile HD).

Individuals affected with HD have a repeat that expanded to 36 or more CAG repeats. It can range as high as 150 or more repeats though most HD alleles have 40 to 50 CAG repeats. These HD alleles can also be grouped in two categories based on their penetrance, i.e. the proportion of individuals who carry the expanded HTT CAG repeat and express the disease phenotype. While HTT alleles with 40 or more CAG repeats are fully penetrant and lead to the onset of HD symptoms, CAG repeats between 35 to 39 repeats are associated with reduced penetrance (Figure 1). Even though some individuals with reduced penetrance alleles will develop HD within their lifetime, typically with a later age-at-onset (AO) and slower progression of the disease, some will never develop clinical manifestations [64-66]. Notably, there have been some reports that suggest that high-normal HTT alleles, with 34 or fewer repeats, can also cause symptoms of HD [67-70], perhaps indicating that these alleles have very dramatically reduced HD penetrance.

CAG repeat length and age-at-onset

The extent to which the HD mutation shows its phenotypic expression (i.e. expressivity) is variable, as individuals with the HD mutation can exhibit differences in many aspects of
their phenotypes, such as their AO of the disease symptoms. While the mutant gene is present from conception and expressed throughout life, in most carriers produces clinical symptoms only after several decades, though these individuals can become symptomatic at any time in their life. The onset of the disease, AO, refers to the age when mutation carriers show the first characteristic signs of HD, typically motor signs such as chorea.

**Figure 2 – The correlation between age at motor onset and the size of the expanded HTT CAG repeat.** Even though the expanded CAG allele is present and widely expressed from conception, the onset of neurological symptoms is inversely correlated with the HTT CAG repeat size (the solid line represents the best-fitting line), such that the longer the repeat the earlier the symptoms manifest. However, individuals with the same repeat size may differ dramatically in AO, for example, individuals carrying a CAG repeat of 40 units can start showing motor signs of HD as earlier as 30 and as late as 90 years of age (the shadow area represents the range of AO for any given CAG repeat size with dashed lines representing the bottom and upper limits). This deviation in AO might be partially explained by the sum of small effects from common genetic variants or by rare variants of large effect, which may act in some but not all individuals (represented by the arrows).

Immediately after the discovery of the HTT CAG repeat, it was shown in different studies that AO was inversely correlated with the size of the expanded CAG repeat [71-73]. The longer the CAG repeat length the earlier the onset of symptoms tends to be, with expansions above 60 CAG repeats usually associated with juvenile HD (Figure 2). Despite the strong correlation between CAG repeat length and AO, it only explains about 50-70% of the variance in motor AO [71-73]. Furthermore, the negative correlation is weaker for repeats at the lower end of the expanded alleles and the severity of the disease does not increase as prominently in the high-expanded alleles where repeat increments do not seem to have as big an effect on phenotype expression [74,75].
CAG repeat instability

While most normal size CAG repeats are stably inherited, expanded CAG repeats are prone to further length variation during meiosis. The pathological alleles are unstable when transmitted to offspring in around 70-80% of transmissions, usually contracting 1 to 2 units if transmitted maternally and expanding 1 to 4 units if transmitted paternally [72,76-83]. Large increases of the repeat size, associated to juvenile HD cases, occur predominantly with paternal transmission, reflecting a particularly high mutation rate during spermatogenesis [78,83-86]. In extremely rare occasions, these larger-sized increases also occur in transmissions of maternal origin [87-90]. Several studies have shown that high-normal alleles, with repeats between 27 to 35 CAGs, can also be unstable and in some cases actually expand into the pathogenic range upon transmission [91-93]. The expanded HD CAG repeat is also somatically unstable, undergoing progressive length increases over time. It is tissue-specific, with the highest levels detected in brain and sperm [80,94].

The origin of Huntington’s disease

A disease-causing mutation may arise on a single chromosome. The haplotype carrying this novel mutation then rises to high frequency by a combination of genetic drift, selection, recombination and population migration resulting in the genetic variation that makes it possible to further infer about the past population dynamics of the mutant allele [reviewed in 95].

The evolution of the HD allele was initially investigated in a simulation study that used several human populations and a variety of primates [63]. These simulations suggested that the human HD allele expanded from a shorter ancestral primate allele, predicting a length dependent mutational bias towards longer HD alleles and a consequent ever-increasing incidence of HD. Indeed, analyses of the mutational flow (defined as the net number mutating past a particular repeat length) have estimated that the new mutation rate in each generation is at least 10% [96]. This is consistent with the proportion of new mutations observed in the University of British Columbia’s DNA and Tissue Bank for Huntington Disease Research Databank, where 14% of alleles with 27-35 CAGs have expanded into the disease associated range [93].
Prevalence of Huntington’s disease

Individuals with HD can be found in many different countries and ethnic groups around the world. However, its prevalence varies geographically, with the highest rates reported for western populations from Europe.

Figure 3 – Estimated prevalence of HD around the world. The worldwide prevalence of HD, based on a meta-analysis study, is estimated at 2.71 per 100,000 [98]. Notably, the largest and best-studied cohort of individuals with HD is in the Lake Maracaibo region of Venezuela, which has the highest reported worldwide prevalence at 700 cases per 100,000 people [101]. The minimum prevalence of HD for each country was based on several epidemiological studies reported in the literature [59,99,102-166].

Europe has a relative high prevalence of HD with approximately 5-10 individuals affected with HD per 100,000 (Figure 3) [97,98]. The prevalence is rather uniform across almost all Europe, with the exception of Finland with a reported prevalence of 0.5 per 100,000 [99]. It is believed that migration out of Europe in the 18th and 19th centuries has brought HD to North and South America, Australia and New Zealand, as prevalence of HD in these regions is similar to the reported in most European countries (5-10 per 100,000) [97,98]. There are, however, some small and isolate communities with an unusual high frequency of HD. In Tasmania, HD was found in one large family of English origin, where 765 descendants of a single ancestor with HD from Somerset (England) were “at-risk” of HD [100]. In Venezuela, the isolated population of Lake Maracaibo has an exceedingly high prevalence rate of up to 700 HD cases per 100,000 individuals [101]. It is believed that HD was introduced in this region between 1860 and 1870 by a Spanish sailor [101].

In contrast to Europe, the HD prevalence rates in Asian populations are significantly lower, with 0.4 and 0.1-0.7 individuals affected with HD per 100,000 in China and Japan, respectively (Figure 3) [108,109,120-122]. The frequency of HD is also expected to be low among people of African origin, with the prevalence in black South Africans estimated at
0.01 per 100,000 [131]. Within the white and mixed-race populations of South Africa the frequency is slightly higher (2.22 and 2.17 per 100,000, respectively) [131], with a founder effect supposedly linking this two populations [167]. Most of these South African HD families are descendants of European settlers, and the HD mutation was in fact traced to a common ancestor of Dutch origin in the Caucasians families [131].

All these prevalence estimates of HD (Figure3) are most likely gross underestimates, as most of these epidemiological data were ascertained before the identification of the HD mutation, and also because nowadays there are still approximately 85% of persons “at-risk” for HD that do not come forward for testing [61].

Haplotypes studies

The geographical distribution of HD suggests that there have been multiple origins of the HD allele with separate new mutations taken place in Europe, Japan and Africa. Indeed, this hypothesis is supported by numerous haplotype studies of a few polymorphisms in the HTT region that have been constructed in different worldwide ethnic populations [62,110,157,167-179].

Many of these haplotype studies have found positive association between specific markers (mostly with the CCG repeat immediately adjacent to the CAG repeat and the deletion of a glutamic acid at position 2642, Δ2642), disease chromosomes and HD prevalence. Interestingly, populations with a high prevalence of HD (e.g. Europeans) have a relatively high number of alleles with the (CCG)7 and Δ2642 (7-del) haplotype, which is associated with a higher number of CAG repeats [62,169]. In contrast, in populations with a low prevalence (e.g. China, Finland, Japan) the 10-in haplotype (with (CCG)10 and no glutamic acid deletion), which is associated with smaller CAG repeats, is most frequent [62,169]. In these populations, the (CCG)7 allele is relatively underrepresented and the Δ2642 allele is rarely found in African and Asian chromosomes [169,177].

More recently, it was identified a subset of 22 tagging single nucleotide polymorphisms (SNPs) in the HTT gene that were highly associated with HD chromosomes from Canadian individuals of European origin. In this population, three major haplotypes (A, B and C) were found in the general population, while HD and high-normal HD chromosomes were almost exclusively found on haplotype A (respectively, 95% and 86% of the chromosomes in these ranges) [180]. In contrast, HD chromosomes in East Asia (China and Japan) were most frequently associated with haplotype C, and none was on haplotype variants A1 or A2, which were the most common in HD chromosomes and present in up to 20% of the general population in Europe [181]. Actually, these two
haplotype variants were completely absent in the East Asia general population sample. The authors suggested that the lower prevalence of HD in East Asia could therefore be explained by the absence of these predisposing haplotype variants that may have arisen in Europe following the separation of the European and Asian populations about 25,000 years ago [181]. On the other hand, HD in the different African subpopulations (black, white and mixed) occurred on diverse and ethnically distinct haplotypes. While HD chromosomes from white and mixed HD patients were predominantly associated with haplotype A, suggesting a similar European origin, the CAG expanded alleles in black South Africans occurred predominantly on haplotype B and C, suggesting multiple origins of the mutation in this ethnic group [182]. Similar to what was previously observed for East Asian populations, there was an absence of the European HD variants in the black South African population, which also may account for the low HD prevalence observed in this population.

The heterogeneous geographic distribution and genetic background of the HD alleles among European, Asian and African population supports the hypothesis that the HD cases spread throughout the world had multiple and separate origins. Furthermore, characterization of these different HD genetic backgrounds may lead to the identification of cis-modifiers affecting HTT repeat stability and pathogenesis and may also provide suitable targets for allele-specific therapeutics.

Genetic modifiers

Even though there is a strong correlation between AO of motor symptoms and CAG repeat size, patients with precisely the same length of the expanded CAG repeat may differ dramatically in AO, approximately ±20 years around the expected AO based on CAG repeat length (Figure 2) [reviewed in 183]. This deviation in motor AO, also known as residual AO, is highly heritable strongly implying the existence of other factors that modulate the rate of the pathogenic process that leads to onset of symptoms [184-186]. The disease expression is likely determined by the combination of multiple stochastic, environmental and/or genetic factors. While all three factors can be involved in the disease manifestation and its timing, genetic modifiers can be more easily identified using genetic techniques directly in studies of human patients, such as candidate gene association studies and, more recently, unbiased scans of chromosomal regions (linkage analysis) and all genome (genome-wide association analysis). By definition, these genetic modifiers of AO are genes whose natural polymorphic variation is capable of modifying the course of the disease pathogenesis, thereby capable of delaying or anticipating the onset of the symptoms. These genes are expected to target particular biochemical pathways
involved in HD pathogenesis, providing new targets for the development of rational therapeutic interventions.

**Candidate gene association studies**

Several polymorphisms in various candidate genes, including within the HTT gene itself, have already been tested in HD patients for an effect of genotype on onset of HD symptoms, mostly on the motor AO. These genes were chosen mainly because they are thought to be functionally relevant to HD pathogenesis or to interact with the mutant huntingtin protein.

**Cis-modifiers**

Within the HTT gene itself, the variation in length of the normal CAG repeat has been extensively tested for genetic association with the AO (Table 1). Kehoe et al. reported in a cohort of 138 Welsh HD families, an effect of the normal repeat length on AO in maternally inherited HD and in male HD patients [187]. A subsequent study in a combined cohort of 754 HD patients from North America and Europe, found evidence of the interaction of

<table>
<thead>
<tr>
<th>Polymorphism</th>
<th>Population</th>
<th>Number of patients</th>
<th>Associated with AO</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (CAG)n</td>
<td>English</td>
<td>293</td>
<td>no</td>
<td>[193]</td>
</tr>
<tr>
<td></td>
<td>Welsh</td>
<td>138</td>
<td>yes</td>
<td>[187]</td>
</tr>
<tr>
<td></td>
<td>Indian</td>
<td>77</td>
<td>no</td>
<td>[190]</td>
</tr>
<tr>
<td></td>
<td>North American/European a</td>
<td>754</td>
<td>yes</td>
<td>[188]</td>
</tr>
<tr>
<td></td>
<td>Venezuelan</td>
<td>443</td>
<td>no</td>
<td>[194]</td>
</tr>
<tr>
<td></td>
<td>European b</td>
<td>921</td>
<td>yes</td>
<td>[189]</td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>404</td>
<td>no</td>
<td>[195]</td>
</tr>
<tr>
<td></td>
<td>Czech</td>
<td>468</td>
<td>no</td>
<td>[196]</td>
</tr>
<tr>
<td>(CCG)n</td>
<td>Italian</td>
<td>80</td>
<td>no</td>
<td>[197]</td>
</tr>
<tr>
<td></td>
<td>French</td>
<td>84</td>
<td>no</td>
<td>[192]</td>
</tr>
<tr>
<td></td>
<td>Indian</td>
<td>77</td>
<td>yes</td>
<td>[190]</td>
</tr>
<tr>
<td></td>
<td>Taiwanese</td>
<td>53</td>
<td>no</td>
<td>[179]</td>
</tr>
<tr>
<td></td>
<td>Venezuelan</td>
<td>306</td>
<td>no</td>
<td>[194]</td>
</tr>
<tr>
<td>Δ2642</td>
<td>Italian</td>
<td>70</td>
<td>no</td>
<td>[197]</td>
</tr>
<tr>
<td></td>
<td>French</td>
<td>79</td>
<td>yes</td>
<td>[191]</td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>293</td>
<td>no</td>
<td>[193]</td>
</tr>
<tr>
<td></td>
<td>French</td>
<td>84</td>
<td>yes</td>
<td>[192]</td>
</tr>
<tr>
<td></td>
<td>Indian</td>
<td>77</td>
<td>no</td>
<td>[190]</td>
</tr>
<tr>
<td></td>
<td>North American/European a</td>
<td>518</td>
<td>no</td>
<td>[184]</td>
</tr>
<tr>
<td></td>
<td>Venezuelan</td>
<td>375</td>
<td>no</td>
<td>[194]</td>
</tr>
</tbody>
</table>

* HD-MAPS cohort; †EHDN Registry cohort
the mutant and normal CAG repeat alleles influencing AO, where the normal CAG repeat was positively associated with delayed onset in HD subjects having large expanded repeat sizes [188]. In addition, a study in a cohort of 921 HD European patients has shown that mutant and normal \( HTT \) alleles interact to influence not only AO, but also clinical severity and disease progression [189]. Besides the normal CAG repeat size two other polymorphisms within the \( HTT \) gene have been implicated, the polymorphic \((\text{CCG})_n\) repeat and the \(\Delta 2642\) (Table 1) [190-192]. Despite these findings, various studies failed to demonstrate any influence of the variation in the normal allele and these polymorphisms on the AO [179,184,190,192-197].

**Trans-modifiers**

Several studies have reported an association of AO with genetic variants within at least twenty-one genes (Table 2). These findings appear to implicate relevant processes such as neurotransmission (\( GRIK2, GRIN2A, GRIN2B, ADORA2\)), energy metabolism (\( PPARGC1A, CO1, NRF-1\) and \(TFAM\)), stress response and apoptosis (\( DFFB, MAP2K6, ASK1\)), gene transcription (\( TCERG1, TP53\)), lipoprotein metabolism (\( APOE\)), autophagy (\( Atg7\)), axonal trafficking (\( HAP1\)), folate metabolism (\( MTHFR\)), DNA repair (\( OGG1\)) and protein degradation (\( UCHL1\)) in the modulation of the pathogenic process that lead to onset of HD symptoms [184,187,190,193-195,198-222].

Although all these candidate genes gave a statistically significant association with AO in the original population studied, many failed when replicated (\( GRIK2, GRIN2B, CO1, PPARGC1A, DFFB, APOE, OGG1, UCHL1, BDNF, TP53\) and \( MTHFR\)) [193,194,199,211,223-231] while some have not yet been confirmed in other populations (\( ASK1, MAP2K6, Atg7, MSX1, NRF-1\) and \(TFAM\)) [184,207,212,216]. Replicable associations have been established for polymorphisms in the \( ADORA2, GRIN2A, TCERG1\) and \( HAP1\) genes [190,194,198,199,204-206,214,217], with the latter presenting the only functional variant defined as a modifier of HD pathogenesis. Note that, in addition to all the genes presented in Table 2, several other biological compelling genes were tested for association but failed to give positive results in the original population (\( GRIN1, GRIN2C, GRIN2D, GSTO1, GSTO2, HIP14, PS-1, TBP, KALRN\)) [204,215,230,232,233].

From the long list of genes in Table 2, those involved in neurotransmission and energy metabolism seem to be the most promising modifier candidates as variants within these genes were replicated in independent cohorts and both pathways were already shown to be compromised in both HD mouse models and patients [reviewed in 234].
One of the proposed mechanisms of HD pathogenesis is the loss of MSNs mediated by the release of glutamate and consequent overstimulation of glutamate receptors causing neurodegeneration and neural damage [235]. Indeed, the first reported candidate genetic modifier was the glutamate receptor ionotopic kainate 2 (GRIK2 also known as GluR6) [193]. Subsequent multiple studies have shown that a rare TAA allele or a functional polymorphism nearby in linkage disequilibrium appears to be associated with earlier onset of HD [200-203]. More recently, polymorphisms within two other genes (GRIN2A and GRIN2B) encoding subunits of the N-methyl-D-aspartate (NMDA) glutamate receptors, known to play a prominent role in synaptic plasticity and neurodegeneration, have shown strong evidences of association with AO [194,204-206].

Abnormal energy metabolism and mitochondrial dysfunction have also been implicated as potential mechanisms involved in the HD pathogenesis [reviewed in 236]. Expression of mutant, but not wild type, huntingtin down-regulates the expression of the mitochondrial regulator peroxisome proliferator-activated receptor γ coactivator 1α (PPARGC1A, also known as PGC-1α) [237-239], a transcriptional regulator of key energetic metabolic pathways, including adaptive thermogenesis [240] and mitochondrial respiration and oxidative stress [241,242]. PGC-1α knock-out mice showed hyperkinetic signs associated with striatal lesions [238] and the crossbreed of these knock-out with HD knock-in mice leads to an increased neurodegeneration of striatal neurons and motor abnormalities [237]. Following these biological findings, three recent independent association studies, including European HD patients mainly from Italy and Germany, pointed PPARGC1A as a modifier of AO of HD symptoms, with the minor allele of rs7665166 associated with a delay in disease onset [208-210].

Even though the candidate association studies has been the favored strategy in the HD field, this candidate gene approach is per se an inherently biased approach in which genes are chosen based on a pre-existing hypothesis related to their connection to pathophysiological processes thought to be involved in the HD pathogenesis.

**Unbiased scans for genetic modifiers**

Another approach that has begun to be employed in HD is the unbiased genetic screen for genetic modifiers, driven by modern technologies that allow a rapid assessment of the overall human genetic variation in a large number of human patients. Genome-wide association (GWA) studies, examining common genetic variants (SNPs) in thousands of HD subjects to see if any variant is associated with age at neurological onset, are now being applied in an expanded version of the HD-MAPS collaboration [185]. These studies
<table>
<thead>
<tr>
<th>Candidate Gene</th>
<th>Polymorphism</th>
<th>Population</th>
<th>Number of patients</th>
<th>Associated with AO</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ADOR2A</strong></td>
<td>rs5751876</td>
<td>French</td>
<td>791</td>
<td>yes</td>
<td>[198]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>German</td>
<td>419</td>
<td>yes</td>
<td>[199]</td>
</tr>
<tr>
<td><strong>GRK2</strong></td>
<td>(TAA), a</td>
<td>English</td>
<td>293</td>
<td>yes</td>
<td>[193]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>North American</td>
<td>258</td>
<td>yes</td>
<td>[200]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>France</td>
<td>138</td>
<td>yes</td>
<td>[201]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Indian</td>
<td>77</td>
<td>yes</td>
<td>[190]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Italian</td>
<td>48</td>
<td>yes</td>
<td>[202]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>European</td>
<td>980</td>
<td>no</td>
<td>[230]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>North American</td>
<td>311</td>
<td>yes</td>
<td>[203]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Venezuelan</td>
<td>368</td>
<td>no</td>
<td>[194]</td>
</tr>
<tr>
<td><strong>GRIN2A</strong></td>
<td>rs1969060</td>
<td>German</td>
<td>167</td>
<td>yes</td>
<td>[204]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Venezuelan</td>
<td>421</td>
<td>yes</td>
<td>[194]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>German</td>
<td>250</td>
<td>yes</td>
<td>[205]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>European a</td>
<td>1069</td>
<td>yes</td>
<td>[206]</td>
</tr>
<tr>
<td></td>
<td>rs8057394</td>
<td>German</td>
<td>250</td>
<td>yes</td>
<td>[205]</td>
</tr>
<tr>
<td></td>
<td>rs2650427</td>
<td>German</td>
<td>250</td>
<td>yes</td>
<td>[205]</td>
</tr>
<tr>
<td><strong>GRIN2B</strong></td>
<td>rs1806201</td>
<td>German</td>
<td>167</td>
<td>yes</td>
<td>[204]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Venezuelan</td>
<td>337</td>
<td>no</td>
<td>[194]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>German</td>
<td>250</td>
<td>yes</td>
<td>[205]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>European b</td>
<td>538</td>
<td>yes d</td>
<td>[206]</td>
</tr>
<tr>
<td></td>
<td>rs890</td>
<td>German</td>
<td>167</td>
<td>yes</td>
<td>[204]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Venezuelan</td>
<td>305</td>
<td>no</td>
<td>[194]</td>
</tr>
<tr>
<td><strong>CO1</strong></td>
<td>C7028T</td>
<td>Italian</td>
<td>51</td>
<td>no</td>
<td>[223]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>German</td>
<td>404</td>
<td>yes</td>
<td>[195]</td>
</tr>
<tr>
<td><strong>NRF-1</strong></td>
<td>rs7781972</td>
<td>German</td>
<td>401</td>
<td>yes</td>
<td>[207]</td>
</tr>
<tr>
<td></td>
<td>rs949152</td>
<td>German</td>
<td>401</td>
<td>yes</td>
<td>[207]</td>
</tr>
<tr>
<td><strong>PPARGC1A</strong></td>
<td>rs7665116</td>
<td>German</td>
<td>400</td>
<td>yes</td>
<td>[208]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Italian</td>
<td>449</td>
<td>yes</td>
<td>[209]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>European</td>
<td>854</td>
<td>yes</td>
<td>[210]</td>
</tr>
<tr>
<td></td>
<td>rs2970870</td>
<td>German</td>
<td>1706</td>
<td>no</td>
<td>[211]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>European b</td>
<td>400</td>
<td>yes</td>
<td>[208]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>European c</td>
<td>854</td>
<td>yes</td>
<td>[210]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>European d</td>
<td>1706</td>
<td>yes</td>
<td>[211]</td>
</tr>
<tr>
<td></td>
<td>rs6821591</td>
<td>Italian</td>
<td>449</td>
<td>yes</td>
<td>[209]</td>
</tr>
<tr>
<td><strong>TFAM</strong></td>
<td>rs1049432</td>
<td>German</td>
<td>401</td>
<td>yes</td>
<td>[207]</td>
</tr>
<tr>
<td></td>
<td>rs11006132</td>
<td>German</td>
<td>401</td>
<td>yes</td>
<td>[207]</td>
</tr>
<tr>
<td><strong>ASK1</strong></td>
<td>rs5880308</td>
<td>German</td>
<td>250</td>
<td>yes</td>
<td>[212]</td>
</tr>
<tr>
<td><strong>DFFB</strong></td>
<td>R196K</td>
<td>Indian</td>
<td>77</td>
<td>yes</td>
<td>[213]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>German</td>
<td>167</td>
<td>no</td>
<td>[224]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Venezuelan</td>
<td>429</td>
<td>no</td>
<td>[194]</td>
</tr>
<tr>
<td><strong>MAP2K6</strong></td>
<td>rs2521354</td>
<td>German</td>
<td>250</td>
<td>yes</td>
<td>[212]</td>
</tr>
<tr>
<td><strong>TCERG1</strong></td>
<td>(Gln-Ala), a</td>
<td>North American</td>
<td>432</td>
<td>yes</td>
<td>[214]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Indian</td>
<td>77</td>
<td>yes</td>
<td>[190]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Venezuelan</td>
<td>427</td>
<td>yes f</td>
<td>[194]</td>
</tr>
<tr>
<td><strong>TP53</strong></td>
<td>R72P</td>
<td>Indian</td>
<td>77</td>
<td>yes</td>
<td>[213]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>German</td>
<td>167</td>
<td>no</td>
<td>[224]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Venezuelan</td>
<td>429</td>
<td>no</td>
<td>[194]</td>
</tr>
</tbody>
</table>
should be able to clarify the suspected involvement of the candidate modifiers genes investigated so far (Table 2) but, more importantly, identify new polymorphisms in genes that would have not been considered based on our current understanding of the disease pathogenesis, revealing thereby completely new biological pathways and processes involved in HD.

While GWAs are still undergoing, initial unbiased scans based on genetic linkage have been used to identify chromosomal regions that harbor modifier genes that can explain residual onset of motor HD symptoms. A whole-genome microsatellite scan in a sample of 692 affected HD sibling pairs (the HD-MAPS study) identified six possible genetic linkage regions (2q33, 4p16, 5q31-32, 6p22, 6q23-24 and 18q22) [243]. A follow-up study, where additional 102 sibling pairs were genotyped for these six genomic regions, has provided strong evidence for a quantitative trait loci (QTL) at the human chromosomal region 6q23-24 modifying neurological onset in HD [244]. A subsequent genome-wide linkage scan comprising 4,500 sibships from the large HD Venezuelan kindreds in which the HD gene was originally identified, revealed two novel loci on chromosome 2 (2p25 and 2q35) and two promising loci on chromosome 5 (5p14 and 5q32) that may contain genes that modify age of onset [245]. This study also confirmed the association of the most promising locus...
from the previous genome scans (on chromosome 6q22) with AO. All the chromosomal regions implicated in these studies are large and so far the specific modifier genes responsible for these effects have not been identified.

**The Huntington’s disease model systems**

Since the location of the human *HTT* gene and the identification of the mutation responsible for HD, numerous HD genetic model systems have been generated. These HD models are non-human biologic systems that can model either the HD mutation or a particular set of HD phenotypes, as well as perturb the *HTT locus*, in order to provide experimentally accessible systems to understand the normal and abnormal function of *HTT*, investigate the disease pathogenesis, test therapeutic approaches and explore their translation to HD patients.

In addition to HD cell lines there are also a wide range of HD animal models, including multicellular eukaryotic organism *Dictyostelium discoideum* [246], invertebrate *Drosophila melanogaster* [247], non-mammalian zebrafish *Danio rerio* [248,249], and mammals mice [250-255], rats [256], sheep [257] and non-human primates [258], that have been genetically engineered to model the HD mutation. Several of these established model systems have been generated by using only the amino-terminal fragment of the huntingtin protein perturbing many cellular processes but removing the polyQ tract from its normal protein context.

**The mouse model**

A number of HD mouse models have been developed that vary in repeat length, level of expression of mutant huntingtin protein and context of the mutant allele, whether it is truncated or full-length, human or murine. The currently available mouse models of HD can be separated in two categories, transgenic and knock-in models. The transgenic mouse models were generated by inserting in the mouse either a truncated (e.g. R6/2 and R6/1 models) [250,252] or a full-length copy (BACHD and YACHD models) [251,253] of the human *HTT* gene that carries an expanded CAG repeat tract. In contrast, knock-in mouse models were generated in a way that the elongated CAG tract was introduced into the mouse homologue of the human *HTT* gene (*Htt*), either only the expanded CAG repeat (*Hdh<sup>Q150</sup> model) [259] or a chimeric human/mouse exon 1 carrying both the expanded CAG repeat and the adjacent polyproline region (e.g. *Hdh<sup>Q111</sup>* and *Hdh<sup>Q140</sup>*) [254,255,260].
The context in which the mutant gene is expressed in these mouse models markedly influences the different disease phenotypes observed. While mice expressing amino-terminal fragments of the HTT exon 1 develop an early onset, rapid progression and aggressive disease phenotype that recapitulates aspects of motor defects (dystonic movements, impairment in motor performance and grip strength) and weight loss seen in HD patients, they are not an exact analogue of adult-onset HD [reviewed in 261,262]. Furthermore, these mice have widespread NIs expression in contrast to the selective neuronal loss in the striatum characteristic of HD [263,264]. This suggests that the full-length huntingtin protein is required to confer this specificity as the initial trigger event seems to occur at the level of the full-length protein rather than small fragments which appear only late in the disorder. Indeed, mice expressing full-length mutant protein (either human or chimeric human/mouse) develop this selective neuronal degeneration and have a more prolonged disease course with fewer prominent motor defects [reviewed in 42], providing an exceptional opportunity to study the evolution of pathogenic processes in the context of a chronically progressing disease phenotype, as observed in HD patients.

**Hdh^{Q111}** knock-in mouse model

HD knock-in mice produce a mutant full-length huntingtin protein with an accurate pattern and level of expression that closely matches the genetic basis of HD. These animals demonstrate a temporal series of histological, biochemical and functional changes that are progressively more severe as the polyQ length increases, show dominance over the normal protein and little discernible effect of the second mutant allele in homozygotes.

One of these knock-in models, the Hdh^{Q111}, was created through the precise insertion of a chimeric human/mouse exon 1 containing an expanded CAG repeat to replace the endogenous mouse exon 1 [254,255]. These mice show a normal life span with the development of mild behavioral abnormalities at later ages; by 24 months of age subtle gait deficits are detected by footprint analysis but no deficits in the accelerated rotarod [265]. Even though knock-in mice do not show the extensive neuronal loss characteristic of HD, nuclear accumulation of full-length huntingtin in striatal neurons becomes evident after several weeks of age. In Hdh^{Q111} homozygous mice, the nuclear huntingtin staining evolves from a diffuse pattern (~6 weeks of age), through a punctate distribution (~5 months) and evident NIs (~10 months), ultimately to neuropil aggregate (~17 months) and reactive gliosis (~24 months) [265]. These knock-in mice also recapitulate many features of repeat instability observed in HD patients. Intergenerational CAG repeat length changes occur in up to 70% of transmissions with a male expansion bias, and tissue specific somatic instability, with significant accumulation of expansions in striatum and liver,
becomes apparent by 5 months of age [254,266-268]. This striatal neuropathology varied when examined in different strain backgrounds, with Sv129 knock-in mice presenting reduced levels of somatic instability and a delayed neuropathology when compared to C57BL/6 knock-in mice [268]. The disease process in this knock-in model also features early energy deficits, supported by decreased levels of cAMP and phospho-CREB that could be responsible for the decreased levels of BDNF also observed in these animals [269] and in HD patients [270,271].

These different studies in the HdhQ111 knock-in model revealed subtle early changes in the HD pathogenic process triggered by the mutant huntingtin that eventually lead to long-term neuropathological consequences that parallel disease in man. Similarly to what is observed in humans, there is a striking mouse-to-mouse variation for some of the phenotypes observed in this mouse model, such as neurodegeneration and motor impairment, suggesting an accumulated impact of many different modifiers in the pathologic process in the mouse [265]. Genetic and phenotypic screens in these mouse models can therefore facilitate the identification of modifiers and its underlying pathophysiological mechanism, ultimately suggesting novel candidate genes for follow-up studies in humans.
Major Goals
This doctoral dissertation is focused on genetic modifiers and their impact on the expressivity of HD, a rare neurodegenerative disorder typically considered to be highly penetrant. The major goal was the discovery of genetic factors that critically influence the pathological process that underlies the manifestation and timing of clinical symptoms of HD, as these are of extreme importance for defining the disease mechanisms and unveiling new therapeutic targets to slow the progression or to ameliorate HD symptoms.

**Specific objectives**

1. To examine the frequency and magnitude of CAG repeat instability in large normal, reduced and full penetrance HD alleles drawn from the Portuguese population

2. Assessment of the CAG repeat site itself and other neighboring common genetic variation
   - 2.1. to test association with differences in HD repeat instability or AO of motor symptoms
   - 2.2. to characterize genetic diversity and origins of the HD mutation in the Portuguese population

3. To replicate, in a large collection of HD samples of European origin, the association of candidate variants in the PPARGC1A, GRIN2A and GRIN2B genes and to consider the potential role of functional variants in genes involved in the dopaminergic pathway

4. To evaluate the contribution of expanded polyQ repeats known to cause other neurodegenerative disorders to the etiology of HD

5. To test the potential overlap of the biology that is modulated by HTT CAG repeat with the complex biology that determines other neurological disorders, namely, amyotrophic lateral sclerosis (ALS) and bipolar disease (BP)

6. To validate, using a mouse model, the effect of a previous HD associated linkage region at human chromosome 6q23-24
Results
Huntington’s Disease Repeat Instability

We examined the frequency and magnitude of HTT CAG repeat instability in order to characterize and understand the genomic and clinical context of repeat instability in the Portuguese population and HD families.

We found that while in control chromosomes drawn from the Portuguese population 6% carried large normal alleles and only one out of 1,772 had a reduced penetrance allele, about 12% of consultants who went through HD pre-symptomatic testing carried either large normal (7%) or reduced penetrance (5%) alleles [Article 1]. Since these ranges of alleles are relatively frequent in the Portuguese population, it is important for genetic counseling of affected families to understand the extent of their instability. Therefore, we traced 13 parent – offspring transmissions of alleles sized 27 to 39 CAGs in 12 families [Article 1]. All 10 large normal alleles were stably transmitted, while one of three (7.7%) reduced penetrance alleles showed instability; specifically, a 39 CAG allele contracted to 38 units when transmitted from a father to his son. Seven small full penetrance alleles (40 to 42 repeats) were seen to contract into the reduced penetrance range, with most (five) being maternal transmissions [Article 1].

Intergenerational instability was also assessed in 134 transmissions from 100 Portuguese HD families. In this study we confirmed a major influence of the gender of the transmitting parent and of the original size of the CAG tract in HD repeat instability, but not the gender of offspring in our population [Article 2]. We reported two large CAG expansions from large repeats, supporting a high intergenerational instability of these repeats even if maternally transmitted; and also one of the largest expansions that resulted in an allele class change, from reduced into full penetrance [Article 2].
Article 1: Large normal and reduced penetrance alleles in Huntington disease: instability in families and frequency at the laboratory, at the clinic and in the population.


Reprinted with the permission of John Wiley and Sons
License number 3190401054225
Short Report

Large normal and reduced penetrance alleles in Huntington disease: instability in families and frequency at the laboratory, at the clinic and in the population

Sequeiros J, Ramos EM, Cerqueira J, Costa MC, Sousa A, Pinto-Basto J, Alonso I. Large normal and reduced penetrance alleles in Huntington disease: instability in families and frequency at the laboratory, at the clinic and in the population.

Large normal (‘intermediate’) alleles may produce de novo expansions in Huntington disease; nevertheless, there is very little evidence about their population prevalence and impact in daily practice, and there are conflicting reports about the extent of their instability. We estimated the frequency of large normal alleles (27–35 CAGs) and of reduced penetrance alleles (36–39 CAGs), as well as the frequency of genotypes carrying them, in (i) a diagnostic laboratory, (ii) a genetic counselling clinic and (iii) the general population. Large normal alleles were present in 6% of a large control sample, 7% of consultands who took pre-symptomatic testing and 7% of samples in the laboratory. Reduced penetrance alleles were found in 1 of 1772 control chromosomes (0.1% of individuals), 5% of 146 pre-symptomatic testees and over 2% of 1214 diagnostic samples (350 families). All 16 alleles sized 27–32 CAGs seemed to be transmitted stably; alleles ≥36 repeats were unstable in five families. Seven small full penetrance alleles contracted into the reduced penetrance range, but none into the large normal range. Evidence showed that large normal alleles are relatively frequent and that those with reduced penetrance are not a rare event, either at the laboratory or the clinic. This reinforces the need to understand the genomic context of repeat instability in each family and population.

Huntington disease (HD) is an autosomal dominant disorder of usually adult onset, characterized by involuntary movements, motor impairment, behavioural changes and cognitive loss. It is caused by the expansion of an unstable (CAG)$_n$ in the first exon of the HD gene (4p16.3), resulting in an extended polyglutamine tract in huntingtin (1). Alleles with 26 CAG repeats or less are normal (class 1), whereas those with 27 to 35 CAGs are classified as ‘intermediate’, large normal, high normal, mutable normal or normal unstable (class 2) and they do not produce HD symptoms. Alleles with 36 to 39 repeats (class 3) show reduced penetrance and are often associated with a later onset, and those with 40 CAGs or more (class 4) are fully penetrant (2, 3).

There is a wide range of age at onset (AO) in HD, inversely correlated with the CAG repeat
Sequeiros et al.

size (4, 5). Prediction models of AO have been refined and are getting more meaningful, particularly for some repeat sizes (6, 7).

The expanded repeat may enlarge further and produce anticipation of AO in successive generations (8, 9) or, less frequently, contract. On occasion, parental transmission may alter the penetrance class of unstable alleles. Although class 1 alleles are stably inherited, those with full penetrance are unstable in about 70–83% of transmissions (10).

Large normal alleles may give rise to new HD cases upon transmission (11, 12), but there are conflicting reports about their degree of instability. In the Venezuelan kindred, only 0.06% showed instability (±1 repeat) and none resulted in a de novo expansion (≥36 CAGs), whereas 14% of reduced penetrance alleles expanded into the full penetrance range (13). In contrast, Semaka et al. reported instability in 30% of transmissions of ‘intermediate’ alleles, 14% of them expanding into the full penetrance range (14).

We have previously reported a relatively high frequency of 27–35 CAG alleles at our diagnostic laboratory, specifically 3.7% of all normal alleles, whereas alleles sized 36–39 CAGs were 3.7% of expanded alleles (3). Later, we reported a frequency of 3% of 27–35 CAG alleles, as well as one 36 CAG allele and one full expansion, among 1772 control chromosomes (15).

Our aim now was to confirm these frequencies and study variation in size upon transmission of large normal and reduced penetrance alleles. This is important for genetic counselling of affected families and to inform the current review process of the international guidelines for the ‘predictive test’ in HD (16).

Materials and methods

CGPP is the reference laboratory for HD testing in Portugal (3). We receive biological samples from physicians all over the country for confirmation or exclusion of their clinical diagnosis and for pre-symptomatic and prenatal testing when there is a family history of HD. The information provided with the sample, including AO and family history, is often limited; when relevant, we have tried to obtain additional information from the requesting physician. Thus, 350 apparently independent kindreds, of variable size, have been tested over the last 12 years. An outpatient clinic is also conducted at CGPP for genetic counselling and pre-symptomatic testing, mostly for families in the northern region, since 1998. Separate consent is obtained for diagnosis and HD research.

Genomic DNA was extracted from peripheral blood lymphocytes by standard methods (17). The (CAG)$_n$ length was determined after polymerized chain reaction (PCR) using fluorescently labelled primers described previously (18); fragments were analysed on an ABI PRISM 3130xl DNA sequencer (Applied Biosystems, Foster City, CA) with GENEMAPPER v. 4.0. Instability was defined as any change in size (≥1 unit) and estimated by subtracting the parental (CAG)$_n$ size from that in offspring (highest peaks). As described previously (15), DNA was also extracted from 1000 anonymous Guthrie cards obtained from the neonatal screening programme, which covers 98–99% of our population; 50 cards from each of the 20 Portuguese districts were randomly sampled.

Results

Control sample (Guthrie cards)

Of the 1000 controls, 886 were successfully typed for their HD (CAG)$_n$ size (Table 1). Large normal alleles were 3% (n = 53) and reduced penetrance alleles were 0.05% (n = 1) of all 1772 chromosomes studied (15). Genotypes with large normal alleles were 6%. There was one genotype with a 36 CAG allele (0.1%), as well as one with a 40 CAG allele (not shown).

Genetic counselling clinic

From January 1998 to June 2009, 146 consultands underwent pre-symptomatic testing (Table 1). Approximately 12% had a class 2 or 3 allele, 6.8% carried a large normal and 4.8% a reduced penetrance allele. In one family, two offspring of an affected mother each inherited one of the two general population large normal alleles present in their father (Fig. 1).

Diagnostic reference laboratory

During the same time period, 1214 tests were performed in HD families (Table 1). Among all alleles typed, 3.5% (n = 85) were large normal (4.8% of all normal alleles) and 1.2% (n = 28) were of reduced penetrance (4.3% of all expansions). Furthermore, 6.8% of the genotypes had at least one large normal allele, whereas 2.4% had at least one reduced penetrance allele (Table 2).

Parent–offspring transmissions

We were able to trace 13 parent–offspring transmissions of alleles sized 27–39 CAGs in 12
Frequency and instability of large normal alleles in HD

Table 1. Distribution of large normal and reduced penetrance alleles in a control sample, the pre-symptomatic genetics clinic and the diagnostic laboratory

<table>
<thead>
<tr>
<th>(CAG)$_n$ size</th>
<th>Control sample$^a$ (886 controls; 1772 alleles)</th>
<th>Pre-symptomatic genetics clinic (146 consultands; 292 alleles)</th>
<th>Diagnostic laboratory (1214 samples; 2428 alleles)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Alleles (%)</td>
<td>Genotypes (%)</td>
<td>Alleles (%)</td>
</tr>
<tr>
<td>27</td>
<td>13 (0.7)</td>
<td>13 (1.5)</td>
<td>3 (1.0)</td>
</tr>
<tr>
<td>28</td>
<td>16 (0.9)</td>
<td>16 (1.8)</td>
<td>—</td>
</tr>
<tr>
<td>29</td>
<td>7 (0.4)</td>
<td>7 (0.8)</td>
<td>—</td>
</tr>
<tr>
<td>Large normal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>6 (0.3)</td>
<td>6 (0.7)</td>
<td>3 (1.0)</td>
</tr>
<tr>
<td>(class 2) alleles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>5 (0.3)</td>
<td>5 (0.6)</td>
<td>2 (0.7)</td>
</tr>
<tr>
<td>32</td>
<td>4 (0.2)</td>
<td>4 (0.5)</td>
<td>2 (0.7)</td>
</tr>
<tr>
<td>33</td>
<td>1 (0.05)</td>
<td>1 (0.1)</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>34</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>35</td>
<td>1 (0.05)</td>
<td>1 (0.1)</td>
<td>—</td>
</tr>
<tr>
<td>Total</td>
<td>53 (3.0)$^c$</td>
<td>53 (6.0)$^d$</td>
<td>11 (3.8)$^a$</td>
</tr>
<tr>
<td>Reduced penetrance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>1 (0.05)</td>
<td>1 (0.1)</td>
<td>3 (1.0)</td>
</tr>
<tr>
<td>(class 3) alleles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>38</td>
<td>—</td>
<td>—</td>
<td>3 (1.0)</td>
</tr>
<tr>
<td>39</td>
<td>—</td>
<td>—</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td>Total</td>
<td>1 (0.05)</td>
<td>1 (0.1)</td>
<td>7 (2.4)$^f$</td>
</tr>
</tbody>
</table>

$^a$Among all 1772 control sample alleles, 1717 were small normal (<26 CAGs) and 1 was a 40 CAG allele.
$^b$Three genotypes were compound (27/28, 27/29 and 30/32 CAGs).
$^c$One genotype was a 30/32 CAG compound.
$^d$All these genotypes had only one class 2 allele and their percentages are, thus, the same as those of the alleles.
$^e$These 11 class 2 alleles were 5.1% of all normal (classes 1 and 2) alleles in the consultands.
$^f$These 85 class 2 alleles were 4.8% of all normal alleles in the laboratory samples.
$^g$These seven class 3 alleles were 9.2% of all expanded (classes 3 and 4) alleles in the consultands.
$^h$These 28 class 3 alleles were 4.3% of all expanded (classes 3 and 4) alleles in the lab samples.

Fig. 1. Pedigree of family UHD377 and a partner with two population large normal alleles uncovered after presymptomatic testing of his offspring. Symbols with a dot represent asymptomatic expansion carriers; current age (years) is shown above each symbol.

families. All 10 large normal alleles were stably transmitted. One of three (7.7%) reduced penetrance alleles showed instability: specifically, a 39 CAG allele contracted to 38 units when transmitted from a father to his son (Table 3).

Table 2. Distribution of genotypes at the diagnostic laboratory, according to allele penetrance classes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Diagnostic laboratory (1214 genotypes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal stable/noral stable</td>
<td>507 41.8%</td>
</tr>
<tr>
<td>Normal stable/large normal</td>
<td>52 4.3%</td>
</tr>
<tr>
<td>Normal stable/reduced penetrance</td>
<td>25 2.1%</td>
</tr>
<tr>
<td>Normal stable/full penetrance</td>
<td>596 49.0%</td>
</tr>
<tr>
<td>Large normal/large normal</td>
<td>3 0.2%</td>
</tr>
<tr>
<td>Large normal/reduced penetrance</td>
<td>— —</td>
</tr>
<tr>
<td>Large normal/full penetrance</td>
<td>27 2.2%</td>
</tr>
<tr>
<td>Reduced penetrance/reduced penetrance</td>
<td>— —</td>
</tr>
<tr>
<td>Reduced penetrance/full penetrance</td>
<td>3 0.2%</td>
</tr>
<tr>
<td>Full penetrance/full penetrance</td>
<td>2 0.2%</td>
</tr>
</tbody>
</table>

Instability in other pairs of relatives

In seven additional families, parents were not typed, but (CAG)$_n$ size was available in sibs and other close relatives (Table 4). All six alleles (three pairs with 27 or 28 CAGs) appear to have been stably inherited. All reduced penetrance alleles (four families) varied consistently in size.

The highest size difference among any relatives was of 8 units and seen in two sibs born 11 years
Sequeiros et al.

Table 3. Transmission of large normal and reduced penetrance HD alleles (13 parent–offspring transmissions in 12 families\textsuperscript{a})

<table>
<thead>
<tr>
<th>(CAG)\textsubscript{n}</th>
<th>Transmitting parent</th>
<th>Stable transmissions ((n = 12))</th>
<th>Unstable transmissions ((n = 1))</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>Two fathers</td>
<td>2 (15.4%)</td>
<td>-</td>
</tr>
<tr>
<td>28</td>
<td>Two mothers</td>
<td>2 (15.4%)</td>
<td>-</td>
</tr>
<tr>
<td>30\textsuperscript{a}</td>
<td>One mother and one father</td>
<td>2 (15.4%)</td>
<td>-</td>
</tr>
<tr>
<td>31</td>
<td>One father</td>
<td>1 (7.7%)</td>
<td>-</td>
</tr>
<tr>
<td>32\textsuperscript{b}</td>
<td>One mother and two fathers</td>
<td>3 (23.0%)</td>
<td>-</td>
</tr>
<tr>
<td>36</td>
<td>One father</td>
<td>1 (7.7%)</td>
<td>-</td>
</tr>
<tr>
<td>39</td>
<td>Two fathers</td>
<td>1 (7.7%)</td>
<td>1\textsuperscript{b} (7.7%)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}One father was a 32/30 CAGs compound (kindred UHD377).
\textsuperscript{b}This paternal allele contracted to a 38 CAG allele in the son.

Results

- Neither parent was affected nor had been tested (Fig. 2a). In a sibship of four, the repeat inherited from their father varied with birth order, from 36 to 41 CAGs (Fig. 2b).

Contraction of full expansions into the reduced penetrance range

- Seven small full penetrance alleles (40–42 repeats) were seen to contract into the reduced penetrance range (Table 5). Most (five) were maternal transmissions.

Discussion

Large normal and reduced penetrance alleles are not rare

- Class 2 (large normal) alleles were present in 6% of controls from our general population, and in 7% of persons at-risk and in 7% of all persons tested at our centre. They represented about 5% of all normal alleles at the diagnostic laboratory.

- Class 3 alleles (reduced penetrance) were seen in nearly 5% of the consultands and in 2.4% of all samples at the diagnostic laboratory, although they were present in only 0.1% of the control sample. They represent 4.3% of all expanded alleles typed at our centre and over 9% of all the expansions found in pre-symptomatic testing.

Class 2 and 3 alleles represent a significant proportion of all genotypes encountered upon genetic testing for HD. Large normal and reduced penetrance alleles should thus be addressed in pre-test counselling, as a possible result. If not previously aware of those potential outcomes, consultands may be very distressed to learn that their test results do not provide a definite prognosis for themselves and/or their children, but rather fall within a grey zone. Also, given the bordering size ranges of the HD repeat and the implications of their exact sizing for diagnosis and for prognosis of AO, alleles at the borderlines between (CAG)\textsubscript{n} size classes should be included as controls for molecular genetics diagnosis of HD.

Large normal alleles seemed to be consistently stable

- We saw no instability at all (\(\geq 1\) unit) in any of 10 observed and 8 inferred transmissions of large normal class 2 alleles (Tables 3 and 4). In the Venezuelan kindred, only 4/69 such alleles changed in size, and none into the pathogenic range (13). However, in the British Columbia kindreds, 33% of large normal alleles were unstable upon transmission, 14% (\(n = 25\)) expanding into the pathogenic range (\(\geq 36\) CAG) resulting in a new mutation for HD (14).

- Sperm analyses showed that class 2 alleles from HD families (‘new mutation intermediate alleles’) are more unstable than those of identical size found in the general population: although assessed only in four males, their risk of expansion into the pathogenic range was 10%, whereas it was 6% for ‘general population intermediate alleles’ (19). However, it has also been postulated that the probability of a male with a large normal allele having an offspring with HD is under 1 in 1000 (20).

Table 4. Large normal alleles show stability among other relatives, whereas those with reduced penetrance were very prone to instability

<table>
<thead>
<tr>
<th>Pedigree</th>
<th>Kinship and transmitting parent</th>
<th>Relative 1</th>
<th>Relative 2</th>
<th>Relative 3</th>
<th>Relative 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>UHD245</td>
<td>Sibs (affected father)</td>
<td>27</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UHD287</td>
<td>Sibs (affected father)</td>
<td>27</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UHD414</td>
<td>First cousins (mother and father are sibs)</td>
<td>28</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UHD444</td>
<td>Sibs (transmitting parent unknown)</td>
<td>44</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UHD134</td>
<td>Sibs (affected father)</td>
<td>36</td>
<td>38</td>
<td>39</td>
<td>41</td>
</tr>
<tr>
<td>UHD201</td>
<td>Sibs (affected father)</td>
<td>39</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UHD214</td>
<td>Paternal uncle – nephew</td>
<td>39</td>
<td>41</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}Relatives are indicated by generation or birth order within generation.

384
Frequency and instability of large normal alleles in HD

Fig. 2. Pedigrees of two families with reduced and full penetrance alleles in the same sibship. Symbols with a dot represent asymptomatic expansion carriers; current age (years) or age-at-death is shown above each symbol. (a) Family UHD044 showing the highest \((\text{CAG})_n\) difference among expanded alleles in relatives (8 units). (b) Family UHD134 showing a progressive increase in \((\text{CAG})_n\) size in successive sibs, offspring of an affected father and a stably transmitted 36 CAG allele.

<table>
<thead>
<tr>
<th>Transferring parent</th>
<th>Parent (CAG)(_n)</th>
<th>Offspring (CAG)(_n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Father</td>
<td>40</td>
<td>38</td>
</tr>
<tr>
<td>Father</td>
<td>40</td>
<td>39</td>
</tr>
<tr>
<td>Mother</td>
<td>40</td>
<td>38</td>
</tr>
<tr>
<td>Mother</td>
<td>41</td>
<td>38</td>
</tr>
<tr>
<td>Mother</td>
<td>41</td>
<td>39</td>
</tr>
<tr>
<td>Mother</td>
<td>41</td>
<td>39</td>
</tr>
<tr>
<td>Mother</td>
<td>42</td>
<td>38</td>
</tr>
</tbody>
</table>

The differences reported in the instability of large normal alleles may be due to ascertainment and sampling methods, diverse population backgrounds and various ancestral haplotypes (14).

Intergenerational instability of reduced penetrance alleles

As in our small series (Tables 3 and 4), several studies have shown reduced penetrance alleles to be highly unstable upon transmission, mainly through the male line, producing full penetrance...
range expansions (11). In the Venezuelan kindred, 14% (6/44) of class 3 alleles expanded further into that range (13). Sperm analyses have shown a mutation frequency as high as 53% for a 36 CAG allele (21).

Reduced penetrance alleles showed instability in one of three parent–offspring pairs (Table 3) and in relatives of all four additional families (Table 4).

Origin of reduced penetrance alleles

Of notice, all seven reduced penetrance alleles, the origin of which we could trace within the family (Table 5), resulted from the maternal or paternal contraction of 40–42 CAG expansions; this is also of great relevance for pre-symptomatic and prenatal counselling. In the Venezuelan kindred, at least 12 of 534 full expansions contracted into class 3 alleles (13). There is also a report of a large contraction, through the maternal line, from 48 CAG repeats into a large normal allele with 34 CAGs (22).

Factors associated with instability of large normal and reduced penetrance alleles

In one sibship (Fig. 2b), there was a progressive increase in repeat size with birth order from reduced to full penetrance alleles. This could be related to an increased instability with the paternal age; however, the father’s (CAG)_n had not been sized.

Previous studies found that large normal alleles lengthening into the pathogenic range had a common haplotype with expanded alleles, defined as haplogroup A by Warby et al. (23, 24). In our sample from the general population, analysis of the (CAG)_n, located next to the (CAG)_n, showed also all three large normal alleles to be associated with a (CCG)_l haplotype, which was seen with 89% of full expansions, but only 60% of class 1 alleles (15).

Further studies are needed

Defining the risk of instability for each (CAG)_n size, within the large normal and reduced penetrance ranges, would fill an important knowledge gap and have great impact in genetic counselling. Large datasets should now be available at diagnostic laboratories and research centers in many countries. Collaborative family and sperm studies of class 2 and 3 alleles should help identifying factors of instability, such as gender and age of the transmitting parent, birth order, population and haplotype background, or other familial factors, in addition to the original repeat size.

Acknowledgements

We would like to thank all patients and all neurologists who collaborated for this study. E. M. R. is the recipient of a scholarship from FCT (SFRH/BD/44335/2008). The second meeting of the genetic testing group of the EHHDN (European Huntington Disease Network) (www.euro-hd.net) and the revision of the IHA ‘predictive testing’ guidelines provided the drive and opportunity for this review of our patient material.

References

Article 2: Intergenerational instability in Huntington disease: extreme repeat changes among 134 transmissions.

Ramos EM, Cerqueira J, Lemos C, Pinto-Basto J, Alonso I, Sequeiros J.


Reprinted with the permission of John Wiley and Sons
License number 3190390285073
Intergenerational Instability in Huntington Disease: Extreme Repeat Changes Among 134 Transmissions

Huntington disease (HD) is a neurodegenerative disorder characterized by motor impairment, chorea, behavioral disturbances, and dementia. It is caused by the expansion of an unstable (CAG)n in the huntingtin gene. While most normal CAG repeats are stably inherited, pathological alleles are unstable when transmitted to offspring, with a strong parental gender effect. The study of intergenerational transmissions may shed light into the factors that influence repeat instability, improve parental counseling and highlight the mechanisms underlying de novo mutations.

Intergenerational instability was assessed in 67 paternal and 67 maternal transmissions, from 100 Portuguese families. CAG repeat length was determined by polymerase chain reaction (PCR) analysis, using fluorescently labeled primers. Instability was estimated by subtracting the parental (CAG)n size from the offspring repeat length.

Instability occurred in 66% of all transmissions and its frequency was similar for paternal and maternal transmissions (Table 1). Male transmission differed markedly from female transmission for mean repeat-length change (P = .003), and for the proportion of expansions, contractions, and stable transmissions (P < .001). Repeat-length variation was strongly related to the initial parental CAG repeat size (P < .001). The change, however, was not correlated with the age of the parent at conception or the gender of the offspring, contrary to what was found in the Venezuelan pedigree. Our results suggest that some factors may be population specific in accordance with the differences observed in the instability upon transmission in different patient cohorts. Also, some ascertainment bias can explain the findings in the Venezuelan pedigree.

Interestingly, the largest increase had a maternal origin. A girl with onset at 8 years of age inherited a large expansion (98 CAGs) from her mother (73 CAGs) who had symptoms since the age of 18 years. This supports a high intergenerational instability of large repeats, even if of maternal origin, which seems to correlate with age-at-onset. These mothers tend to have their children at a very young age (<25 years). There is also a large difference between the 2 siblings (unle and mother of the girl with 98 CAGs) in this family (brother with 48 and sister with 73 CAGs), who inherited the HD CAG expanded allele from their father, showing very high instability in only 2 generations.

The largest paternal CAG expansion observed was from 1 father (32 CAGs) who transmitted a 76 and 69 CAG allele to his 2 daughters. Among the paternal transmissions we have also observed an interesting case from a family without previous history of HD, in which a 48 CAG repeat was detected in a male with depression since the age of 20 years and chorea at the age of 40 years. His father has a 36 CAG allele and no disease symptoms at the age of 72 years. This large paternal expansion (+12) is 1 of the largest expansions involving a reduced penetrance allele and resulted in allele class change into full penetrance.

Our study confirms a major influence of the gender of the transmitting parent and of the original size of the CAG tract in HD repeat instability, but not the gender of offspring in our population. Also, our reported families with extreme changes in repeat lengths upon transmission suggest the existence of major family-specific instability modifiers that may be highly heritable. These extreme repeat-length changes have implications in genetic counseling and, in the context of prenatal diagnosis, special attention should be taken in young females with large repeats and in carriers of reduced penetrance alleles.


Eliana Marisa Ramos, MSc, Joana Cerqueira, MSc, Carolina Lemos, PhD, Jorge Pinto-Basto, MD, Isabel Alonso, PhD, and Jorge Sequeiros, MD, PhD.

Unigen (Unit for Genetic and Epidemiological Research in Neurological Diseases), IBMC - Institute for Molecular and Cell Biology, Universidade do Porto, Porto, Portugal; 2CGPP (Centre for Predictive and Preventive Genetics), IBMC - Institute for Molecular and Cell Biology, Universidade do Porto, Porto, Portugal; 3ICBAS (Instituto de Ciências Biomédicas Abel Salazar), Universidade do Porto, Porto, Portugal.

*Correspondence to: Dr. Jorge Sequeiros, Unigen, and CGPP, IBMC, Universidade do Porto, Rua Campo Alegre 823, 4150-180 Porto, Portugal; e-mail: js@ibmc.up.pt.

Funding agencies: Financiamento Plurianual de Unidades de Investigação de FCT (Fundação para a Ciência e a Tecnologia). E.M.R. is the recipient of a scholarship from FCT (SFRH/BD/44335/2008). L.A. is funded by Programa Ciência 2005. P.S.P.H (Programa Operacional do Potencial Humano) - QREN (Quadro de Referência Estratégico Nacional) - Programa 4.2 - Promoção do Emprego Científico, and is co-funded by ESF (European Science Foundation) and national funds by the Ministério da Ciência e Ensino Superior (MCTES).

Relevant conflicts of interest/financial disclosures: Nothing to report. Full financial disclosures and author roles may be found in the online version of this article.

Published online in Wiley Online Library (wileyonlinelibrary.com), DOI: 10.1002/mds.24065
### Results – Article 2

#### Table 1. Transmission of the expanded HD repeat in Portuguese families

<table>
<thead>
<tr>
<th></th>
<th>Paternal</th>
<th>Maternal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Daughters</td>
</tr>
<tr>
<td>Transmissions, n</td>
<td>134</td>
<td>67</td>
</tr>
<tr>
<td>(CAG), in parent (mean ± SD)</td>
<td>42.95 ± 4.33</td>
<td>42.15 ± 2.82</td>
</tr>
<tr>
<td>Mean age at conception (mean ± SD)</td>
<td>27.96 ± 5.94</td>
<td>28.82 ± 5.33</td>
</tr>
<tr>
<td>Unstable transmissions, n (%)</td>
<td>80 (66)</td>
<td>46 (69)</td>
</tr>
<tr>
<td>Contractions, n (%)</td>
<td>33 (25)</td>
<td>8 (12)</td>
</tr>
<tr>
<td>Expansions, n (%)</td>
<td>55 (41)</td>
<td>38 (57)</td>
</tr>
<tr>
<td>Repeat-length variation (mean ± SD)</td>
<td>+1.12 ± 3.91</td>
<td>+2.10 ± 4.19</td>
</tr>
<tr>
<td>Contractions (mean ± SD)</td>
<td>−1.52 ± 0.71</td>
<td>−1.56 ± 0.52</td>
</tr>
<tr>
<td>Expansions (mean ± SD)</td>
<td>+3.64 ± 5.08</td>
<td>+4.00 ± 4.73</td>
</tr>
<tr>
<td>Repeat-length variation (range)</td>
<td>−4 to +25</td>
<td>−2 to +24</td>
</tr>
</tbody>
</table>

HD, Huntington’s disease; SD, standard deviation.

#### References


**Huntington’s Disease cis-modifiers**

**Huntington’s Disease normal allele**

Even though it is well known that HD is caused by an expanded *HTT* CAG repeat and that the length of this repeat is strongly correlated with the onset of symptoms, there has been much debate in the HD field whether variation within the normal *HTT* CAG repeat can modulate HD pathogenesis. Given the critical implications of such assumptions, we tested, using the largest dataset so far, whether the normal allele, the interaction between the expanded and normal alleles, or the presence of a second expanded allele could influence AO of HD motor signs.

Our study showed that HD pathogenesis leading to motor symptoms is determined by a completely dominant action of the longest expanded allele and that there is clearly no additive effect from the second mutant allele in HD “homozygotes” subjects [Article 3]. Moreover, our data uncovered the profound effect that even one single CAG outlier sample, with a very long CAG repeat and extremely young AO relative to all others, can have on the final result when testing for the effects of individual modifier genes. HD cases with CAG alleles in the 40 to 53 repeat range, where shown here to yield a well-behaved data set that conforms to the fundamental assumptions of linear regression analysis (constant variance and normally distributed error) [Article 3].
**Article 3:** CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion.


CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion

ABSTRACT

Objective: Age at onset of diagnostic motor manifestations in Huntington disease (HD) is strongly correlated with an expanded CAG trinucleotide repeat. The length of the normal CAG repeat allele has been reported also to influence age at onset, in interaction with the expanded allele. Due to profound implications for disease mechanism and modification, we tested whether the normal allele, interaction between the expanded and normal alleles, or presence of a second expanded allele affects age at onset of HD motor signs.

Methods: We modeled natural log-transformed age at onset as a function of CAG repeat lengths of expanded and normal alleles and their interaction by linear regression.

Results: An apparently significant effect of interaction on age at motor onset among 4,068 subjects was dependent on a single outlier data point. A rigorous statistical analysis with a well-behaved dataset that conformed to the fundamental assumptions of linear regression (e.g., constant variance and normally distributed error) revealed significance only for the expanded CAG repeat, with no effect of the normal CAG repeat. Ten subjects with 2 expanded alleles showed an age at motor onset consistent with the length of the longer expanded allele.

Conclusions: Normal allele CAG length, interaction between expanded and normal alleles, and presence of a second expanded allele do not influence age at onset of motor manifestations, indicating that the rate of HD pathogenesis leading to motor diagnosis is determined by a completely dominant action of the longest expanded allele and as yet unidentified genetic or environmental factors. Neurology 2012;78:690-695

GLOSSARY

HD = Huntington disease.

Huntington disease (HD) is a neurodegenerative disorder with motor impairment, cognitive decline, and psychiatric manifestations caused by an expanded CAG trinucleotide repeat (>35 CAGs) in the gene encoding huntingtin. The expansion shows a very strong negative correlation with age at onset of motor signs. Although the dominant inheritance pattern of HD indicates that one expanded allele is sufficient to trigger the disorder, polymorphism of the normal allele CAG repeat (i.e., the number of consecutive CAGs in the nonexpanded allele; <36 CAGs) has been suggested in some but not all studies to influence the timing of onset of disease manifestations. Recently, an interaction of the expanded and normal allele was reported in a large study to modify age at onset based upon motor signs, cognitive change, and behavioral manifestations. Counterintuitively, longer normal alleles (e.g., 30 CAGs) seemingly delayed age at onset of subjects with longer expanded CAG alleles (e.g., >50 CAGs). This finding would have important implications for the precise molecular mechanism that initiates HD pathogenesis, as it would suggest that individual huntingtin molecules might physically interact, with a resultant alteration of the pathogenic potential of the expanded repeat. Similarly, if the normal allele can modify HD pathogenesis, it would provide a potential...
route to therapeutic intervention to delay or prevent onset via a genetic modifier validated to act via a mechanism that operates in humans with HD. Consequently, we re-explored this critical issue using a much larger dataset. Specifically, we aimed to test whether the normal allele CAG repeat, either alone or in interaction with the expanded allele, or the presence of a second expanded allele modifies age at onset of HD motor signs.

METHODS Subjects. We analysed DNA of 4,078 patients with HD (1,865 men, 1,845 women, and 389 unreported) with known age at onset of overt motor manifestations, including 10 subjects with 2 expanded HD alleles ("HD homozygote"). DNA samples were from subjects involved in long-term genetic studies in the Massachusetts HD Center Without Walk, either from the local region or from collaborating investigators (HD-MAPS), and from 3 large HD observational studies (COHORT, PREDICT-HD, and REGISTRY). The HD-CAG repeat length was determined using a modification of the PCR amplification assay reported by Warner et al., with a fluorescent oligonucleotide primer pair flanking the repeat for automated allele calling after capillary electrophoresis on an ABI3730XL DNA Analyzer, using a set of HD CAG allele standards determined by DNA sequencing, as reported.1 The means of mutant and normal CAG repeat lengths of HD homozygote subjects (4,068) were 44.86 (range, 36–120; median, 44) and 18.48 (range, 9–35; median, 18), respectively (figure e-1 on the Neurology® Web site at www.neurology.org).

Standard protocol approvals, registrations, and patient consents. This study used only deidentified, previously collected DNA samples and phenotypic data in a manner approved by the Institutional Review Board of Partners HealthCare, Inc.

Statistical analysis. Primarily, this study utilized samples from 4,068 HD homozygote subjects with 1 expanded HD allele. Natural log-transformed age at onset of motor signs was modeled as a function of mutant HD CAG repeat length, normal HD CAG repeat length, and interaction of mutant and normal allele as independent variables using a linear regression analysis. For models involving interaction terms, all continuous variables were centered around the mean value of each variable. Initially, statistical models were fitted to data including all subjects. Subsequently, as described in the supplemental data, we built models using only normally distributed data points in the restricted range of 40 to 53 CAG repeats, where there is no evidence of heteroscedasticity. All statistical analyses were performed using R (version 2.7.2).

RESULTS Regression analysis of HD heterozygotes. For all subjects with one expanded HD allele (4,068; figure e-1), the relationship with age at motor onset is shown for both the expanded and normal allele CAG repeat lengths in figure 1, A and B, respectively. For this dataset, we modeled natural log-transformed age at onset by 1) expanded allele CAG repeat length (greater than 35 repeats), 2) normal allele CAG repeat length (35 or fewer repeats), and 3) interaction between them using a linear regression analysis, as has been the common practice in investigation of allele length effects in CAG repeat disorders. Although we initially observed an apparent significant association for the interaction of the normal and mutant alleles, we discovered that this result (table 1) relied on a single outlier with a mutant allele of 120 CAGs and a normal allele of 11 CAGs (shown in blue in figure 1, A and B). That one individual sample could have such a profound effect on the final result indicated that in testing for genetic modifiers,
the analysis of HD age-at-onset data requires a more detailed, rigorous statistical analysis that better conforms with the fundamental assumptions of linear regression analysis (normally distributed data points with constant variance). Consequently, we performed a series of statistical analyses to explore this question (figures e-2, e-3, e-4, and e-5), including analysis of subjects with expanded CAG repeats between 40 and 53, with a formal exclusion of outliers, and a separate analysis of the outlier individuals excluded from the model. In no case did we obtain a statistically significant result supporting either an independent effect of the normal CAG repeat or an interaction of the normal and mutant CAG repeats (table 1; table e-1). We also tested directly whether subjects with shorter normal allele CAG repeat lengths had a different age at onset than subjects with longer normal allele CAG repeat lengths (figure e-6), again finding no significant effect of the normal CAG allele on age at onset. Thus, our results indicate that the expanded allele CAG repeat length is the single most important determinant of the age at onset of motor signs of HD and that the normal CAG allele length does not play a significant modifying role in the pathogenic process leading to motor onset.

HD “heterozygotes.” Importantly, in addition to the heterozygote subjects analyzed above, there were 10 subjects with 2 expanded alleles in our dataset. Although the sample size is small, these subjects would have the potential to reveal any major additive or synergistic effects of possessing 2 expanded alleles and no normal allele. There is clearly no additive effect, as all of these individuals would be expected to show motor onset at <4 years of age if this were determined by the sum of the 2 expanded CAG repeats. Interestingly, subjects with 2 expanded alleles (figure 1C; red, longer allele; green, shorter allele) did not show a significant deviation from the minimal adequate statistical model (from figure e-5 data) when plotted based upon the longer of their 2 expanded alleles (figure 1C, red line). The ages at onset of these subjects resided within 2 standard deviations of each allele-specific age at onset spectrum. Combined with our findings from HD heterozygotes, these data indicate that age at onset of motor manifestations in HD is primarily determined by a single allelic dose of the mutant gene and is not influenced dramatically by the length of the normal allele, by the presence of a second mutant allele, or by the absence of a normal allele.

**DISCUSSION** It is well known that HD is initiated by a dominant action of an expanded CAG repeat in the HD gene and that age at onset strongly correlates with expanded allele CAG repeat length. However, there has been much debate in the field whether age at onset is influenced by the second HD allele, especially the length of the CAG repeat on the normal chromosome.\(^6\)\(^-\)\(^8\)\(^-\)\(^12\)\(^-\)\(^13\) This is a critical question, as the answer determines whether the pathogenic effect of the expanded polyglutamine tract in mutant huntingtin protein can be modulated by expression of an equivalent huntingtin protein with a polyglutamine repeat in the normal size range. A positive finding would clearly have ramifications both for the current effort to develop a therapy by nonselective RNA-mediated suppression of mutant and normal HD alleles and also for the potential of using wild-type huntingtin as a therapeutic protein in HD.

Using the largest dataset so far, with stringent statistical analyses, our results have dismissed a significant role for the normal allele CAG repeat length in modifying age at onset of motor manifestations, indicating that the expanded allele CAG repeat length is not only the initial trigger of HD pathogenesis but also the predominant factor determining the rate of the process that leads to motor onset. We found no evidence that an interaction between the expanded allele and the normal allele is an explanation for any portion of the variance in age at onset of motor signs that is not already explained by the expanded allele alone. Furthermore, our findings with a larger series of CAG repeat genotyped HD heterozygotes support previous reports that, at lengths over 40 repeats (repeats <40 may show nonpenetrance), a single allelic

### Table 1

<table>
<thead>
<tr>
<th>Regression model(^a) (CAG range)</th>
<th>No. of subjects</th>
<th>Intercept (p value)</th>
<th>Expanded CAG allele (p value)</th>
<th>Normal CAG allele (p value)</th>
<th>Interaction (p value)</th>
<th>R(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial model (36-120)</td>
<td>4,008</td>
<td>-0.00023 (0.943)</td>
<td>-0.00256 (-2.10(^{-14}))</td>
<td>0.00106 (0.267)</td>
<td>-0.00006 (0.998)</td>
<td>0.657</td>
</tr>
<tr>
<td>Initial model excluding one 120 CAG subject (36-99)</td>
<td>4,007</td>
<td>-0.000003 (0.999)</td>
<td>-0.00304 (-2.10(^{-14}))</td>
<td>0.000084 (0.376)</td>
<td>-0.00013 (0.506)</td>
<td>0.634</td>
</tr>
</tbody>
</table>

\(^a\) Log (age at onset of motor manifestations) = expanded + normal + expanded - normal, using centered variables.

\(^b\) See figure e-5.
dose of the HD mutation is sufficient to cause HD within a typical human lifespan and to determine age at motor onset.\textsuperscript{16,21} Though the 10 homoyzgotes studied are too few to exclude any onset-modifying effect of the second mutant allele on a statistical basis, any such effect would have to be quite small compared to the effect of the single mutant allele discerned in HD heterozygotes. This suggests that the level of mutant huntingtin protein produced from a single allele already exceeds any minimum threshold required to trigger pathogenesis (at a rate determined primarily by its CAG repeat length) and that neither additional mutant protein nor the absence of any normal protein further alters the rate of pathogenesis leading to motor onset.

Although our results clearly showed no effect of the normal allele CAG repeat length on age at onset, other csi-factors may play a role in modifying age at onset. For example, a gene closely linked to the HD gene has been hypothesized to modify age at onset\textsuperscript{14} and, in the HD-MAPS study, suggestive evidence for linkage was found at chromosome 4p16.\textsuperscript{12} Therefore, it will also be important to determine a potential role for other csi-factors in potentially modifying age at onset of HD manifestations. In addition, although neither the normal HD CAG allele repeat length nor its interaction with the expanded allele influenced age at motor onset, the remaining variance has been reported to be highly heritable,\textsuperscript{15,17} indicating the presence in the genome of genetic modifiers. Such modifier genes would provide potential therapeutic targets already validated to alter the course of HD in human patients and so their identification is a very active area of investigation. The findings reported here provide the basis for more reliable statistical assessment of the role of individual modifier genes and should contribute to their unequivocal identification.

AUTHOR AFFILIATIONS
From the Center for Human Genetic Research (J.-M.L., J.-H.L., T.G., J.S.M., D.L., M.E.M., J.F.G.), Massachusetts General Hospital, Boston; LonGEn, IMBC-Institute for Molecular and Cellular Biology (E.M.R.), and Instituto de Ciencias Biomédicas Abel Salazar (J.S.), Universidade do Porto, Porto, Portugal; Center for Molecular Medicine and Therapeutics (M.R.H., S.C.W.), University of British Columbia, Vancouver, BC, Canada; Regional Medical Genetics Centre, Fairford HSC Trust, Truro, and University of Ulster (P.M.S), Coleraine, UK; Hemepen County Medical Center (M.N.), Minneapolis, MN; Department of Psychiatry and Behavioral Sciences (C.A.R., L.R.M.), Johns Hopkins University, Baltimore, MD; Neurogenetics Unit and Centre for Rare Diseases (J.S., S.O.), IRCCS Neuroted, Pizzitelli, Italy; Istituto Nazionale Neurologico C. Besta (S.D.D.), Milan, Italy; Department of Neurology (E.G.T.), and Department of Genetics (C.A.), Fundação Instituto de Diagnóstico, São Paulo, Brazil; Department of Clinical Neurosciences (O.S.), University of Calgary, Calgary, AB, Canada; Sydney Medical School (R.J.A.T.), University of Sydney, and Department of Neurology (E.M.), Westmead Hospital, Sydney, NSW, Australia; Department of Biology (A.N.), University Tor Vergata, and Institute of Neurobioloy and Molecular Medicine (I.F.), Rome, Italy; Center for Neurodegenerative Disease (R.J.), Emory University School of Medicine, Atlanta, GA; Department of Neurology (T.A.), University of Florida, Gainesville; Department of Neurology (S.J., M.H.S.H., R.H.M.), Boston University School of Medicine, Boston, MA; Max-Planck Institute for Neurodegenerative Disease (S.M.H., H.D.R.), Munich, Germany; and Department of Neurology (M.R.H.), University of Virginia, Charlottesville; Memory and Aging Center (A.Z.), University of California, San Francisco; Department of Neuropsychiatry and Behavioral Science (R.K.A.), University of South Carolina School of Medicine, Columbia; Department of Neurology (K.M.), Columbia University Medical Center, New York, NY; Department of Psychiatry and Neurology (J.S.P.), The University of Iowa Roy and Lucille Carver College of Medicine, Iowa City; Department of Neurology (G.B.L.), Ulm University, Ulm, Germany.

AUTHOR CONTRIBUTIONS
Dr. J.-M. Lee conceived of and performed the statistical analysis and interpretation and drafted the manuscript. Drs. Rams, J.-H. Lee, Gillis, and Myore generated and interpreted molecular data. Drs. Hayden, Wirby, Morrison, Nance, Ross, Morgulis, Spataro, Ondreka, Di Durrans, Gomez-Tomasa, Ayuso, Suchowersky, Trem, McCloud, Novellino, Frontali, Jones, Ahamimak, Frank, Saint-Hilaire, Hendi, Reina, Lucenz, Harrison, Zamb, Aharonson, Minkoff, and Seppälä contributed to the conceptualization and design of the study and revised the manuscripts for intellectual content. Drs. Paulsen and Landenheimer revised the manuscripts for intellectual content. Drs. Myers, Mac-Donald, and Gussola participated in conceptualization and design of the study and interpretation of the data and revised the manuscript for intellectual content.

ACKNOWLEDGMENT
The authors thank the study participants and their families, who made this work possible, coinvestigators and contributors from the HD-MAPS study, the Huntington Study Group (PREDICT-HD and COHORT studies), and the European Huntington’s Disease Network (REGISTRY study (see appendix e-1)) as well as the National Research Roster for Huntington Disease Patients and Families.

STUDY FUNDING
Supported by grants from the National Institute of Neurological Disorders and Stroke (NINDS) (“Huntington’s Disease Center Without Walls”), NS016507, the CHDI Foundation, Inc., and the Huntington’s Disease Society of America’s Coalition for the Cure. J.H.L. received a fellowship from the National Research Foundation of Korea and E.M.R. received a scholarship from the Fundação para a Ciência e Tecnologia of Portugal (FCT). PREDICT-HD is supported by NINDS grant NS040968. Both PREDICT-HD and COHORT are also supported by the CHDI Foundation, Inc. and REGISTRY is supported by the High Q Foundation.

DISCLOSURE
Dr. J.-M. Lee reports no disclosures. Dr. Rams received a scholarship from FCT (SFRH/BD/49252/2008). Dr. J.-H. Lee received support from the National Research Foundation of Korea (NRF-2009-009210). Dr. Gillis and Dr. Myore report no disclosures. Dr. Hayden serves as Editor-in-Chief of Clinical Genetics. Dr. Wirby reports no disclosures. Dr. Morrison is a section editor of The Oncologist. Dr. Nance received honoraria from the Huntington’s Disease Society of America HD Training Course for Physicians, the American Academy of Neurology Genetics in Neurology Course and lectures at Augusta College, royalties from Oxford University Press for Juvenile Huntington’s Disease 2009, reimbursement for NIH Study Section service and the Parkinson Study Group Scientific Review Board, Center of Excellence support from the Huntington Disease Society of America and the National Parkinson Foundation, and grant support from Santhera Pharmaceuticals, NeuroSearch, Medivation/Pfizer, Schwez Biosciences, IMPSAX Pharmaceuticals, Neuraxis Pharmaceuticals and Tama Pharmaceuticals, National Institutes of Health, NINDS grants NS041466, NS06630, NS049068, N001392, and NS064416, NCCAM grant AT002613, and NIMH grant HG010449, and grants from the CHDI Foundation (HP Therapeutics, the High Q Foundation, and the Michael J. Fox Foundation).

Neurology 78 • March 6, 2012
693

Genetic modifiers of Huntington’s disease
47
Northwestern Dixon Foundation, and dedicates annual participation in
Speaker’s bureaus: Roche, Sanofi. Dr. Ross received research funding from
National Institutes of Health NINDS grants NS16375, NS15532, NS080018, and NS069422. Dr. Mangione received re-
search funding from National Institutes of Health NINDS grant NS016375. Medivation/Pfizer, CHDI Foundation Inc., and Century
Laboratories and was a consultant to AstaZeneca, Dr. Squier and
Dr. Ostrerello report no disclosures. Dr. Di Donato was on the Sci-
tific Advisory Board of the Fondazione Miani Arti, Milan, Italy, and a
consultant to the Scientific Advisor for Preclinical Research and Interna-
tional Relationships of the Scientific Director of the Fondazione IRCCS Institute Neurologico Carlo Besta, Milan, Italy, and received support
from the Ministero della Salute, Italy, and from the Fondazione Cariplo,
Milan, Italy. Dr. Gomez-Tortosa reports no disclosures. Dr. Ayuso has
received research support from CIBERER, RCS6, Madrid, Spain. Dr. Szech
wierski was an Advisory Board Member for Allergan, Biowall and
Novartis, on the editorial boards of the Canadian Journal of Neurological Sciences and the Journal of Neurology, and received support from
Abbott Laboratories, from National Institutes of Health NINDS grant
NS044331, and from the CHDI Foundation, Inc. Dr. Tessma reports roy-
talties for the book Molecular Medicine, 3rd edition (Elsevier) and support
from Australian Research Council LEID grants LE159891 and LE16001830. Dr. McCusker is on the editorial board of Trends in
Hepatology and Hepatico-Pancreatic-Metabolism. Dr. Tessma reports no disclosures. Dr. Fran-
tuli received support from Italian Drug Agency (AIFA) grant
FARM63997TX. Dr. Jones received support from the National Institutes of Health NINDS grants NS40868 and NS66918 and from the High Q
Foundation and the Huntington’s Disease Society of America. Dr. Ashizawa received support from the National Institutes of Health NINDS
grant NS568897. Dr. Frank consulted for and received support from Lundbeck Pharmaceuticals. Dr. Saint-Hilaire was on the Steering Com-
mittee of the Clinical Trials (MOTIONS) study of EMED nimodipine. Instead at
the meetings of the Movement Disorder Society and the American Acad-
emy of Neurology, is a member of the speaker’s bureau of Teva Pharma-
caceuticals, and received support from Bayer, EMED Sante, and Teva
Pharmaceuticals. National Institutes of Health NINDS grants NS059995, and
NHGRI grant HG0441 and from the Michael J. Fox Foundation.
Dr. Nechiporuk consults with Merck and Aventis Pharmaceuticals and
received support from the National Institutes of Health NINDS grants
NS40868 and NS66918, and from the High Q Foundation and the Huntington’s Disease Society of America. Dr. Zanetto reports no disclosures. Dr. Abumro received support from National Insti-
tutes of Health NINDS grant HD62556, South Carolina Council on Disability grant 18185.3100, and a grant from the Fullerton Founda-
tion. Dr. Milder served on the editorial board of Neurology® and is an
editor of Current Neurology and Neuroscience and received support from
Amarin Corporation, NeuroSearch Sweden AB, and Medivation, and from
National Institutes of Health NINDS grants NS03630, NCRR
grant RB024156, and PO12116-N from NHGRI grant HG0441 and from the University of Rice and as well as from the Parkinson Disease
Foundation, the Huntington’s Disease Society of America, the Parkinson Study
Group, the Michael J. Fox Foundation, and the CHDI Foundation, Inc.
Dr. Squier has a compensated speaker for the Foundation Jamiens Dyst.
also served on the editorial boards of Clinical Genetics, the Journal of
Community Genetics, and the Journal of Neurology and Biochemistry, and
received research support from the FCT “Financiamento Pluri-
anual de Unidades de Investigação” and from the European Commissi-
on FP6 Network of Excellence EuroGenetic: Genotyping in Europe. Dr. Paulsen received funding from National Institutes of
Health NINDS grant NS040668 and from the CHDI Foundation, Inc.
Dr. Landshamer received a speaker honorarium/reimbursement from
dCGNP-Templier Pharma. royalties from Neurolucida Compact
(Velcro) and Juvenile HD (Oxford Press), consulting support from the
CHDI Foundation, Inc., clinical trial support from Medivation and Neu-
rosearch, and support from the European Union (FP7 Paddleton) and
the High Q Foundation/CHDI Foundation. Dr. Myers received funding
from National Institutes of Health NINDS grant NS016287 and from
the Jerry MacDonald Huntington Research Fund. Dr. MacDonald serves
as an Associate Editor of PLoS Genetics, is on the Board of Directors of the
Huntington’s Disease Society of America, and received support from
National Institutes of Health NINDS grants NS016267, NS023765, and
NS035648 and from the Huntington’s Disease Society of America and
the CHDI Foundation, Inc. Dr. Gosnell is on the Scientific Advisory
Board of Quest Diagnostics, Inc., is on the editorial boards of DNA and
Cell Biology, Neurology of Disease, Neurogenetics, Brain Control Neu-
sciences, Brain Control Biology, and Molecular Autism, and received sup-
port from National Institutes of Health NINDS grants NS016267, NS024279, and NS069422, NHGRI grants GM050534 and NICHD
grant HD06528 and from the Simon’s Foundation Autism Research Ini-
tiative, Autism Speaks, CHDI Foundation, Inc., and the Huntington’s
Disease Society of America.

Received March 13, 2011. Accepted in final form July 18, 2011.

REFERENCES
321.
2. Schoenfeld M, Myers RH, Cappeles LA, Berkman B, Sax
DS, Clark E. Increased rate of suicide among patients with
3. Andrew SE, Goldberg YP, Kremer B, et al. The relation-
ship between trinucleotide (CAG) repeat length and clinical
features of Huntington’s disease. Nat Genet 1993;4:
369–368.
4. Duyao M, Ambrose C, Myers R, et al. Trinucleotide re-
peat length instability and age of onset in Huntington’s
between trinucleotide repeat expansion and phenotypic
370.
disease CAG trinucleotide repeats in pathologically con-
7. Klemper J, Zikovic J, Stocki J, Ing VK, Uhrbo T, Roth J.
The number of CAG repeats within the normal allele does
not influence the age of onset in Huntington’s disease.
Mov Disord 2011;26:125–129.
8. Warner JP, Barron JF, Brock DJ. A new polymerase
chain reaction (PCR) assay for the trinucleotide repeat that
is unstable and expanded on Huntington’s disease chro-
and mutant HTT interact to affect clinical severity and
progression in Huntington disease. Neurology 2009;73:
1280–1285.
complete penetrance Huntington’s disease alleles among
individuals with major depressive disorder. Am J Psychia-
try 2010;167:574–579.
normal and expanded CAG repeat sizes influences age at
**Supplementary data**

**Figure e-1**

![Graph showing distributions of expanded (mutant) and normal HD CAG repeat lengths in subjects used in this study.](image)

*Figure e-1. Distributions of expanded (mutant) and normal HD CAG repeat lengths in subjects used in this study.*

Frequencies of chromosomes for each CAG repeat length of normal (open) and mutant allele (expanded; red) were plotted.
**Figure e-2. The initial model violates statistical assumptions of linear regression**

The traditional linear regression analysis used often in the field yielded a model that violated statistical assumptions of constant variance and normally distributed error, as demonstrated in panels A-C.

(A) To test the assumption of equal variance, we compared residuals calculated from the model to fitted values (predicted centered log onset age).

(B) To assess the normality of the initial model, we compared actual residuals to theoretical residuals from a normal distribution in a quantile-quantile plot.

(C) To identify influential data points in the initial model, leverage and the Cook’s distance were plotted against the residuals. Leverage is commonly used to identify observations that have a large effect on the regression model, and a data point with high leverage indicates that that observation is distantly located from the center of the measurements. Cook’s distance estimates the influence of data points on model fit by measuring the effect of deleting a given observation.

Blue circle in Panels A and C represents the HD subject with an expanded *HD* repeat length of 120 CAG trinucleotides.
Figure e-3. Variance and normality of age at motor onset distribution for HD subjects.
The problems of non-constant variance and non-normal error were evident in the observations of age at onset at individual expanded allele CAG repeat lengths. (A) Variance of age at onset, represented by standard deviation (SD) was not constant. (B) Natural log transformation of the age at onset (LN(onset)) provided a partial remedy for the range of adult-onset associated repeat lengths that form the bulk of the sample. (C) A Box plot of natural log-transformed age at onset against expanded allele showed the existence of outliers (open circles) defined by a standard quartile method (outside of 1.5 times interquartile range from first or third quartile).
Figure e-4. Distributions of age at onset by individual expanded CAG repeat allele.
Exclusion of the outliers identified statistically in Figure e-3 resulted in distributions of log-transformed age at onset that resembled the normal distribution for adult onset-associated CAG repeat lengths: histograms of log-transformed age at onset for each expanded allele CAG repeat length were overlaid by a theoretical normal distribution (red line). Numbers represent expanded allele CAG repeat length. Number at top right represents expanded allele CAG repeat length. LN(onset), natural log transformed age at onset.
Figure e-5. An updated model with a well-behaved subsample of the dataset

To update the model to conform with the assumptions for linear regression, we excluded data points contributing to non-normal error or non-constant variance. We used 3,674 subjects with expanded allele CAG repeat lengths between 40 and 53, after excluding 198 and 70 subjects with higher or lower repeats, respectively and 126 outliers. The final well-behaved subset comprised 90.3% of the original sample.

(A) The relationship between expanded allele CAG repeat length and age at onset of motor signs.

(B) The relationship between normal allele CAG repeat length and age at onset of motor signs.

(C) Residuals calculated from the updated model compared to fitted values.

(D) Comparison of actual residuals to theoretical residuals from a normal distribution in a quantile-quantile plot.

(E) Leverage and the Cook’s distance plotted against the residuals.
Figure e-6. Comparisons between subjects with shorter or longer normal allele CAG repeat lengths. (A) Two linear regression models were generated using subjects with shorter normal allele CAG repeat length (normal allele CAG < 17, blue line and blue circle; 616 subjects) and subjects with longer normal allele CAG repeat length (normal allele CAG > 21, red line and red circle; 576 subjects). Log-transformed age at onset of normally distributed data points was modeled by expanded allele repeat length. Expanded allele CAG repeat lengths between the two groups were not significantly different (Mann-Whitney U test, p value, 0.7336). Distributions of age at onset of subjects with shorter normal allele CAG repeat length (B) and those with longer normal allele CAG repeat length (C) were plotted. The distributions of age at onset were not significantly different between the two groups (Mann-Whitney U test, p value, 0.5488).
Table e-1. Statistical analyses of samples excluded from the updated model as phenotypic or repeat length outliers.

<table>
<thead>
<tr>
<th>Model $^5$ (CAG range)</th>
<th>Number of Subjects</th>
<th>Intercept (p-value)</th>
<th>Normal CAG (p-value)</th>
<th>R-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model using phenotypic outliers (40-53)</td>
<td>126</td>
<td>-0.48792 (0.001)</td>
<td>-0.00151 (0.841)</td>
<td>0.00032</td>
</tr>
<tr>
<td>Model using subjects with CAG $&gt;$ 53 (54-99)</td>
<td>197</td>
<td>0.27252 (0.136)</td>
<td>0.00833 (0.401)</td>
<td>0.00361</td>
</tr>
<tr>
<td>Model using all subjects excluded in the updated model (36-99)</td>
<td>393</td>
<td>0.16350 (0.290)</td>
<td>-0.00800 (0.333)</td>
<td>0.00240</td>
</tr>
</tbody>
</table>

$^5$ Residuals were calculated based on the minimum adequate model: the updated model (without centering) from Figure e-5 using well-behaved data points (CAG, 40-53; 3674 subjects), and then, the residual was modeled as a function of length of the normal CAG allele.
Huntington’s Disease haplotypes

To investigate the origins of the HTT CAG mutation and assess the role of HD cis-elements as modifiers of repeat instability and AO, we performed extensive haplotype studies for the region flanking the HTT gene.

Detailed analysis of genome-wide SNP data from about 700 HD chromosomes of Western/Northern European origin delineated a prominent ancestral haplotype, the most frequent HD haplotype also in other studies, that it is marked by the Δ2642 deletion allele. This haplotype is enriched in high-end normal/reduced-penetrance CAG alleles that exhibit unstable germ-line transmission and can give rise to new cases of HD. However, these HTT haplotypes were not associated with altered CAG-repeat length distribution or AO of motor symptoms, arguing against common cis-regulatory elements modulating HD pathogenesis [Supplementary Article 1].

We have used some of the highly informative SNPs and single tandem repeat (STR) markers spanning a ~1Mb region around HTT to infer the phased haplotype of over 300 unrelated multiplex HD families of Portuguese origin. Our study revealed a variety of haplotypes in this population, where genetic diversity decreased as CAG repeat size increased, with large normal allele chromosomes sharing the same major haplotypes as the HD chromosomes [Article 4]. One of these haplotypes, marked by the HTT Δ2642 deletion, was actually shared by the two biggest HD expansions observed in our set [Article 2], implying some level of historical germ-line instability. To test this hypothesis we genotyped additional Southern-European families from South Tyrolean isolated populations in Italy. We observed the occurrence of intergenerational repeat instability in non-pathological repeat range alleles in the other major haplotypes, arguing against cis-elements mediating CAG repeat instability and de novo expansions [Article 4].
**Article 4:** Haplotype analysis of the HTT 4p16.3 region in the Portuguese population.

**Ramos EM,** Gillis T, Mysore JS, Lee JM, Gögele M, D’Elia Y, Sequeiros J, Pramstaller PP, Gusella JF, MacDonald ME, Alonso I.

[in preparation]
Haplotypet analysis of the HTT 4p16.3 region in the Portuguese population

Eliana Marisa Ramos, Tammy Gillis, Jayalakshmi S. Mysore, Jong-Min Lee, Martin Gögele, Yuri D’Elia, Jorge Sequeiros, Peter P. Pramstaller, James F. Gusella, Marcy E. MacDonald, Isabel Alonso

From the Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA (E.M.R., T.G., J.S.M., J.-M.L., J.F.G., M.E.M.); UniGENe, IBMC – Institute for Molecular and Cell Biology, University of Porto, Porto, Portugal (E.M.R., J.S., I.A.); EURAC- European Academy of Bozen/Bolzano, Bolzano, Italy (M.G., Y.D., P.P.P.); CGPP, IBMC – Institute for Molecular and Cell Biology, University of Porto, Porto, Portugal (J.S., I.A.); ICBAS – Instituto de Ciências Biomédicas Abel Salazar, University of Porto, Porto, Portugal (J.S., I.A.)

Corresponding Author Isabel Alonso, UniGENe and CGPP, IBMC, University of Porto, Rua Campo Alegre 823, 4150-180 Porto, Portugal
Telephone: +351 22 607 4942
Fax: +351 22 600 2923
Email: ialonso@ibmc.up.pt

Running title HTT haplotypes in Portuguese population

Abstract Huntington’s disease (HD) is a dominantly inherited neurodegenerative disorder characterized by involuntary choreic movements, cognitive impairment and behavioral changes. It is caused by the expansion of an unstable CAG repeat in the HTT gene. In this study we characterized the genetic diversity of the HD mutation in the Portuguese population, by performing an extensive haplotype analysis of ~1Mb region flanking the HD gene, in over 300 HD families with Portuguese origin. We observed high haplotype diversity in both normal and expanded chromosomes, representative of more than one ancestral chromosome underlying HD in Portugal, where multiple independent events have given rise to the expansion into HD allele range. One of the haplotypes (core haplotype A) marked by the HTT delta2642 deletion, was enriched in HD chromosomes
and seemed to be particularly prone to expansion (shared by the two largest expansions reported in the Portuguese population). However, a closer look at normal HD alleles with higher CAG repeats revealed a higher similarity in the extended haplotype with HD alleles. We also observed that one of the largest (+12 CAGs) expansions resulting in an allele class change into full penetrance happened on a core haplotype B. These findings suggest that these range of alleles are a reservoir for new HD mutations, independent of the haplotype. To further support this hypothesis, we have also looked at unstable transmissions observed in an isolate population from Italy, and observed that in the non-pathological allele range instability occurs in other haplotypes besides haplotype A.

**Introduction** Huntington’s disease (HD) is a dominantly inherited neurodegenerative disorder, usually of adult onset, characterized by involuntary choreic movements, cognitive impairment and behavioral changes. HD is caused by the expansion of an unstable polymorphic CAG repeat in the *HTT* gene located on chromosome 4p16.3 [1], with alleles between 35-39 repeats showing incomplete penetrance while alleles with 40 or more repeats are fully penetrant. Alleles with fewer than 35 CAGs are considered to produce no symptoms of HD. HD age-at-onset (AO) of clinically diagnosed symptoms is inversely correlated with the number of expanded CAGs; repeat length alone explains about 50-70% of the variance in AO [2], while the remaining variance in AO is highly heritable. Expanded CAG repeats are prone to further length variation, while most normal size CAG repeats are stably inherited. However, there are conflicting reports about the degree of instability of high-normal alleles, a subclass of normal alleles with repeats between 27 and 35 CAGs. Several studies have shown that these alleles can also be unstable [3-5], while in the Venezuelan kindred and Portuguese population these alleles were mostly stably inherited [6, 7].

Several haplotype studies in HD families of different ethnic populations, have investigated the origins of HD by constructing haplotypes that included only a few polymorphisms in the *HTT* region, mainly the adjacent proline repeat and the deletion of a glutamic acid residue at position 2642 (delta2642) [8-13]. More recently, two studies have characterized in more detail the genetic diversity around the *HTT* region. Warby et al. found that 95% of (268 European) disease chromosomes are associated with the same haplotype (haplotype A), which is also significantly enriched (83%) on high-normal alleles (27-35 CAGs) relative to general population [14]. They hypothesized that *cis*-acting factors have a major predisposing influence on HD CAG instability. In a subsequent study, they found that in East Asia (China and Japan) most HD chromosomes are associated with a minor HD European haplotype (haplotype C) [15]. More recently these authors have shown that
HD chromosomes from white and mixed African HD populations were mostly associated with haplotype A, while in black Africans HD occurred predominantly on haplotype B and C [16]. The European HD haplotypes (A1 and A2) were absent in both Asian and African general population. The authors argued that different HTT haplotypes have different mutation rates and that the geographic differences in the haplotypes explain the difference in HD worldwide prevalence. In a different study, Lee et al., identified 40 single nucleotide polymorphisms (SNPs) centered on the HTT gene that showed genome wide-significance for association with HD. Their findings revealed that even though the HTT region exhibits many different haplotypes, one apparent ancestral haplotype, marked by the HTT delta2642 deletion, accounted for about 50% of HD chromosomes and extended to at least 938Kb on about half of these [17]. However in this study, neither the extended shared haplotype nor the individual local HTT haplotype were associated with altered CAG-repeat length distribution or residual age at onset for motor symptoms. The major haplotype identified in this study has been previously associated with sporadic cases of HD in “new mutation” families [4].

A small haplotype study in 140 HD Portuguese families, showed three different HD founder haplotypes associated with 7- (most frequent among normal and HD chromosomes), 9- and 10-CCG repeats, suggesting the possibility of different origins for the HD mutation among the Portuguese population [18]. Following up these findings, and in order to characterize in more detail the genetic diversity and gain more insight on the origins of the HD mutation in the Portuguese population, we performed an extensive haplotype analysis of the region flanking the HD gene, in over 300 HD families with Portuguese origin.

Material and Methods

Portuguese HD families. DNA samples were collected at CGGP, a reference laboratory for HD testing in Portugal, which receives biological samples from physicians all over the country for confirmation/exclusion, pre-symptomatic and prenatal testing for HD. A total of apparently independent 334 Portuguese HD families, consisting of 531 individuals, were used for haplotype studies. It consisted of 441 individuals with an expanded HD allele (sixteen of which had a repeat in the reduced penetrance range), 18 with a high-normal HD allele (27-35 CAGs) and 72 unaffected relatives (<27 CAGs). Informed consent was obtained from all the individuals that participated in this study.

Non-HD families: MICROS study. The MICROS study is a population-based survey carried out in three small isolated villages of the Val Venosta, South Tyrol, in Italy. The villages of Stelvio, Vallefunga and Martello were selected for their geographical and
historical isolation. A detailed description of the MICROS study can be found in Pattaro et al [19]. Briefly, DNA was collected from over 1,400 volunteers and genealogical information was reconstructed, by means of church records and municipality lists, going back to the 1600s (15 generations).

**Selection and markers genotyping.** Selection of the majority of the markers used in this study was based on previous work from our lab that showed that these markers were overrepresented on HD chromosomes [17]. Genotyping of nine SNPs (rs6814736, rs7667745, rs1751848, rs762847, rs7658462, rs3129317, rs1730768, rs12641989 and rs16844364; Figure 1) flanking the HD gene was performed by real-time polymerase chain reaction (PCR) using the commercially available Taqman Genotyping probes (Applied Biosystems, Foster City, CA) carried out on the LightCycler ® 480 (Roche Diagnostics, Mannheim, Germany), following manufacturer’s instructions. Repeat sizes of the HD CAG repeat, proline repeat and D4S127 and genotyping of the in/del polymorphism (delta2642) were determined by using previously established, but slightly modified, PCR amplification assays, using fluorescently labelled primers [20-23]. The size of the fragments was then determined using the ABI PRISM 3730xl automated DNA Sequencer (Applied Biosystems, Foster City, CA) and GeneMapper version 3.7 software. For each marker, a set of sequenced samples was used as standards.

**Haplotype phasing.** PHASE software (http://stephenslab.uchicago.edu/software.html) version 2.1 was used to reconstruct haplotypes from genotypic data when the phase could not be directly inferred by family structure. Haplotypes were reconstructed by incorporating the genotypes of SNPs rs6814736, rs7667745, rs1751848, rs762847, rs7658462, rs31293, rs1730768, rs12641989 and rs16844364, as well as the repeat sizes of the D4S127 and the CCG repeat encoding proline (as a microsatellite locus). The genotypes of the HD CAG repeat mutation (expanded or normal) and the delta2642 polymorphism (insertion or deletion) were also included. Phase-known haplotypes were taken in account and only haplotype pairs with a PHASE probability greater than 0.8 were used for further analyses.

**Statistical analysis.** Since the HTT CAG distribution and repeat-length variation among the different core haplotypes did not display a normal distribution, these two variables were analyzed and compared between haplotypes by using nonparametric tests: the independent samples Kruskal-Wallis test and the independent samples median test. All the statistical analyses were performed using PASW Statistics 18 (SPSS Inc., Chicago, IL).
Results

Extended ~1Mb haplotype shared by 19% of Portuguese HD chromosomes. We extracted genotypes from a region of 1.04 Mb around the HD gene for 805 independent chromosomes, from which we were able to generate 786 haplotypes after phasing, for 317 HD and 469 control chromosomes. In the extended haplotype, which incorporates the entire 1.04Mb region around HTT, we included SNPs rs6814736, rs7667745, rs1751848, rs762847, rs7658462, rs3129317, rs1730768, rs12641989 and rs16844364, the proline repeat and the delta2642 deletion. We were able to construct complete phased haplotypes at all these eleven loci for 193 HD and 267 control chromosomes. The remaining HD and control chromosomes were not considered in this first analysis because some chromosomes were missing genotypes or could not be ambiguously phased at all loci. In the Portuguese HD chromosomes, we found 39 different extended haplotypes, with the top 4 haplotypes (haplotypes 1 to 4 in Figure 1) present in almost 50% of all the 193 HD chromosomes, while only shared by 18% of all the 267 control chromosomes (Supplementary Table 1 and 2). The most common of these extended HD haplotypes, haplotype 1, which was shared by 36 out of 193 (18.7%) HD chromosomes but only in 10 out of 267 (3.8%) control chromosomes, includes the delta2642 deletion allele, shown previously to be enriched in HD chromosomes [8, 13, 21, 24], and the 7-CCG repeat, most frequent in both mutated and normal Western-European alleles [9, 13, 18, 25, 26]. The second and third most common HD haplotypes (haplotype 2 and 3) were, respectively, the third and fourth most frequent haplotypes on the control chromosomes (haplotype 2 accounted for 14.5% of the HD and 7.1% of control chromosomes, while haplotype 3 represented 6.7% of HD and 5.6% of control chromosomes). The other HD major haplotype, haplotype 4, was present in 6.2% of all HD chromosomes and was relatively rare in the control chromosomes (1.87%). Out of the 39 haplotypes found in the HD chromosomes, 21 were not found in the control chromosomes and were present in about ¼ of all HD chromosomes. On the other hand, we found 68 different extended haplotypes in our Portuguese control chromosomes (Supplementary Table 2). From these, only 18 haplotypes were also found on the HD chromosomes. The two most frequent extended haplotypes on the control chromosomes, haplotype 9 (15.4%) and haplotype 20 (7.87%), were shared by only 3.1% and 1.55% of all HD chromosomes. Neither of these haplotypes (9 and 20) included the delta2642 deletion codon allele. Noteworthy, we found a higher genetic diversity among the control chromosomes, where 50 low frequency haplotypes accounted for 45% of normal chromosomes.

Enrichment of core haplotype A among Portuguese HD chromosomes. In order to understand if some of the rarer HD haplotypes could derived from any of the four major
Results – Article 4

extended haplotypes, we grouped the haplotypes based in a smaller region (318Kb) around the HD CAG repeat that included rs762847 (164Kb upstream of the HD CAG repeat), the adjacent CCG proline repeat and the delta2642 polymorphism (154Kb downstream the repeat), as shown in Figure 1. The 317 independent haplotypes found in the HD chromosomes were combined into eight different “core” haplotypes (haplotypes A to H in Figure 1). Haplotypes found in the control chromosomes but not on the HD chromosomes were grouped together and designated “other”. Two major core haplotypes (A and B) accounted for more than 60% of all Portuguese HD chromosomes (Figure 2). Core haplotype A, which included among others the most common extended haplotype (haplotype 1), was present in about ¼ of all HD chromosomes while only present in 6% of the control chromosomes. Notably, almost all the chromosomes (63 out of 68 genotyped for the D4S127 marker) in this core haplotype A included 157-bp allele for the D4S127 marker, a single tandem repeat (STR) previously shown to be highly associated with the HD mutation [27], suggesting that these chromosomes are indeed ancestrally related. Also, we observed that ~80% chromosomes within this core haplotype shared the complete downstream region (up to 362Kb downstream the HD CAG repeat), while there was more diversity in the region upstream the marker rs762847 (Figure 3). Even though core haplotype A included the most common extended haplotype, core haplotype B (without the delta2642 deletion allele) is the most frequent core haplotype and is equally represented among HD (38%) and control (33%) chromosomes. By typing the additional STR marker proximal to the CAG repeat, we observed that most of the HD chromosomes (83 out of 107 genotyped for the D4S127 marker) included in this core haplotype shared the same 155-bp allele, while in the control chromosomes we observed more diversity with three frequent alleles, 153-, 155- and 157bp alleles corresponding to respectively, 24.1%, 28.6% and 42.9% of the control chromosomes genotyped for D4S127 within core haplotype B. This finding reflected the higher genetic diversity observed among the control chromosomes both upstream and downstream of the CAG repeat (Figure 3). On the other hand, one of the most frequent core haplotypes among control chromosomes (27%), core haplotype D, was present only in 11% of the HD chromosomes, and most chromosomes within this core haplotype shared the 151-bp STR allele. Most of the genetic variability in this core haplotype was observed upstream of the HD CAG repeat (Figure 3). HD chromosomes within core haplotype C had a similar frequency to the control chromosomes (15% and 17%, respectively). However, while in the HD chromosomes, the STR revealed the presence of two possible variants within these chromosomes, one with the same 151-bp allele (38%), and other with the 155-bp allele (42%), most of the control chromosomes (54 out of 65 genotyped for the D4S127 marker) had the 151-bp allele.
**High-normal alleles of the major haplotypes genetically closer to HD alleles.** The CAG repeat sizes for all the chromosomes of core haplotypes A to D are shown in Figure 4. We observed that core haplotype A occurs on chromosomes with repeat sizes ranging from 9 to 73 repeats, with a distribution shifted toward higher repeat sizes (median CAG=42.0, IQR 40.0 to 45.0), while in core haplotypes C and D the distribution was shifted toward smaller repeat sizes (median=19.5, IQR 17.0 to 42.0, core haplotype C; median=18.0, IQR 17.0 to 23.0, core haplotype D). Core haplotype B, the most frequent core haplotype in both HD and control chromosomes, ranged from 12 to 87 CAG repeats with a median of 24.5 repeats (IQR 18.0 to 42.0). Both distributions (Independent samples Kruskal Wallis Test, \( p < 0.001 \), and medians Independent samples median test, \( p < 0.001 \)) of the HD CAG repeat were significantly different across the four different core haplotypes. Even considering only chromosomes in the normal HD CAG repeat range (<36 CAG repeats), core haplotype A showed a higher median CAG repeat size than the other 3 major core haplotypes (26.0, core haplotype A vs.18.0, core haplotype B; 17.0, core haplotype C; 17.0, core haplotype D). When comparing the extended haplotype of the control chromosomes with higher CAG repeats (20-35 CAG repeats), we observed that for the two core haplotypes with higher variability among control chromosomes (haplotype B and C), these higher CAG alleles shared a greater percentage of identity with the correspondent extended haplotype (haplotype 2 and 4, respectively). Noticeably, in haplotype B higher CAG chromosomes had a higher percentage of the 155-bp STR allele, present in 81% of all HD chromosomes within this haplotype, while the lower CAG chromosomes had mostly STR alleles 153-bp and 157-bp.

**HD haplotypes and repeat instability.** We were able to collect intergenerational instability data for 83 of the families included in this study and looked at the associated HD haplotype of 41 maternal and 42 paternal transmissions (Table 1). In this subset of HD transmissions, the mean repeat-length variation was +0.59 (±4.14) for maternal and +2.69 (±4.46) for paternal transmissions. For the core haplotype A, we observed a mean repeat-length variation (+2.2 for maternal and +4.9 for paternal transmissions) numerically greater than the overall (and other core haplotypes) for both maternal and paternal transmissions, although there were no significant differences between groups. Furthermore, both the largest maternal (+25) and paternal (+24) transmission observed in Portuguese families [28] shared the same core haplotype A. However, the paternal transmission reported to be one of the largest (+12) expansions involving a reduced penetrance allele (36 CAGs) that resulted in a full penetrance allele (48 CAGs) [28], occurred on a core haplotype B.
Unstable normal alleles and haplotypes. We determined the HD CAG repeat for about 1,300 members of an Italian population isolate [19] with complete genealogy information up to 15 generations. We were able to trace 472 transmissions of the HTT CAG repeat from 120 closely related “nuclear families”. From these, we only found two unstable transmissions of normal range HTT CAG alleles in one “nuclear family” (Figure 5, Family 1). In this consanguineous family (with a common ancestor within 4 meiosis), the paternal 23 CAG repeats allele on haplotype C contracted 1 repeat, while the maternal 20 CAG repeat allele on haplotype B contracted 2 CAG repeats. We also observed an unstable transmission of a high-normal HD CAG repeat (a 32 CAG allele expanded to 33 CAG repeats), which occur on a haplotype C chromosome inherited from the father (Figure 5, Family 2). We were also able to trace 22 transmissions of the HTT CAG repeat, from 13 closely related “nuclear families”, that were associated with the del2642 allele (that marked haplotype A and also the rare haplotypes E and G). The transmitted HTT CAG repeat had mostly repeat sizes between 20 to 24 CAG repeats, while in two sibships was associated with a 12 CAG repeat allele. None of these HTT CAG repeat transmissions associated with the del2642 allele were unstable.

Discussion Originally, it was believed that the HD mutation emerged among Western Europeans [29], mainly due to the highest rates of HD in this population. However, after the discovery of the HD causal mutation, subsequent analysis of the HTT region showed that the geographical variation in prevalence might be correlated with the frequency of certain polymorphic alleles that resulted on a variety of independent and not easily related HD haplotypes, suggesting different origins for the HD mutation [8, 13, 30].

Even though, in our study, we focused on a particular and restricted population from Portugal in the South of Europe, we found 39 different extended haplotypes in the Portuguese HD chromosomes. Despite this diversity, about 19% of all HD chromosomes shared an identical haplotype (haplotype 1) that extended over 680Kb downstream and 362Kb upstream the HD CAG repeat and was marked by the delta2642 deletion allele. This major HD haplotype corresponds to the most frequent HD haplotype defined in earlier studies, and seems to be the same or extremely similar extended haplotype found to account for about 25% of a set of 699 European HD chromosomes [17]. The fact that we can observed such a high proportion of HD chromosomes sharing an identical SNP haplotype in such an extended region and in even apparently closed related populations (Western-European versus Southern-European) but shown to have intra-European variation in allele frequency [31], reflects the low recombination rate across this 1Mb region.
As shown in previous studies [8, 13, 21, 24], we found an increased frequency of the delta2642 deletion allele among the Portuguese HD chromosomes, present in almost 27% of all HD chromosomes whereas only 6% of the control chromosomes had this rare allele. This suggests that some of the 39 HD haplotypes bearing the delta2642 deletion allele might be related and that recombination events lead to a greater genetic diversity on the ~1Mb region around the HD gene. Therefore, we reduced the region of interest to a smaller core of 318Kb around the HD repeat itself. By restricting our haplotype reconstruction to this smaller region, we were able to group related extended haplotypes and retrieve eight different core haplotypes associated with HD chromosomes, with four major haplotypes accounting for almost all HD and control chromosomes. These 4 major core haplotypes are not easily related to each other anymore: haplotype A (enriched in HD) is the only bearing the delta2642 deletion allele, while haplotype D (enriched in controls) is the only with a 10-CCG repeat. Furthermore, by looking at the STR marker upstream the HD repeat, we observed a nonrandom allelic association between the HD chromosomes with core haplotypes (A, B and D) and D4S127. Haplotype A and D are marked, respectively, by the 157-bp and 151-bp allele, in both HD and control chromosomes, while haplotype B is associated mainly with the 155-bp allele. This marker seems to be pretty stable and even though slippage can happen it occurs in a clearly small enough percentage that linkage disequilibrium can still be detected and therefore also the correspondent extended haplotype. These two observations reinforce the view that many of the haplotypes observed in our population, represent indeed independent events that lead to HD mutation.

In a more detailed analysis of the four different haplotypes, we also observed some genetic variability in the region flanking the HTT gene among chromosomes within the same lower resolution haplotype, suggesting that multiple occurrences of the CAG expansion seemed to have happened within the same haplotype. Variants within each refined haplotype are easily related to the correspondent extended haplotype and can be mostly found in the pool of different haplotypes in the normal population, suggesting that rare events and mechanisms like recombination, gene conversion and/or slippage altered the original haplotype found in the normal population, that afterwards due to slippage during DNA replication expanded into the HD repeat range. Our data suggests that this variability may be explained by the occurrence of not only one of these mechanisms but rather, for a given haplotype, the pool of HD chromosomes may have arisen through a combination of various different events. For example, in haplotype C, we found an absence of some of the extended variants found in the HD chromosomes, suggesting that these are not due to independent events but rather to the occurrence of
more than one recombination event (mostly double recombination) in addition to slippage in the STR. We cannot exclude genetic admixture that would also explain the introduction of these new genetic lineages into the Portuguese population.

Early HD studies have suggested that particular haplotypes, mainly marked by the delta2642 deletion allele and/or the 7-CCG allele, seemed to be particularly predisposed to expansion that eventually would result in a repeat allele within the HD range, supported by an enrichment of repeats in the high normal range in these particular haplotypes [8, 9, 13, 18, 32]. We observed that, also in our Portuguese control chromosomes, the haplotype marked by 7-CCG and delta2642 deletion alleles (haplotype A) had a higher median CAG repeat length than any of the other three major haplotypes (with most of the normal alleles ranging from 20 to 33 repeats). As suggested by others, it appears to be HTT haplotypes where the HD mutation is more likely to occur. Even though we cannot ruled out the hypothesis that cis-acting factors have a major predisposing effect on the instability of the HD repeat, it is also possible that haplotype A - which is present in only 6% of the control chromosomes, has a bias toward higher repeat alleles as a result of genetic drift [33]. Repeat-length variation is strongly related to the initial parental CAG size, therefore a larger average repeat size would make haplotype A more likely to expand and become a new HD mutation. Indeed, this specific haplotype seemed to be particularly prone to expansion as it was interestingly shared by the two highest expansions observed in the Portuguese population (a +24 paternal and a +25 maternal transmission). However, we also observed a large paternal expansion on haplotype B (+12 CAGs) that resulted in a repeat class change from reduced to full penetrance range. Furthermore, both haplotypes B and C carried alleles with really large repeats. Actually, the highest repeat size in this study was observed in haplotype B (87 CAG repeats), suggesting that also these haplotypes seem to be prone to big expansions, not found in this particular study probably due to missing information on these pedigrees. To further support this hypothesis, we compared the extended haplotype of high normal CAG repeats (>20 CAGs) in all four major haplotypes and observed that these shared a higher percentage of similarity with HD chromosomes than shorter CAG repeat alleles, suggesting that instead of a predisposing haplotype, there is a pool of high normal alleles in each different founder haplotype more prone to give rise to HD alleles. In an isolate population from Italy [19] we were able to trace 472 transmissions of HTT CAG repeats in the normal range, and only in three instances the repeat was unstably transmitted. In concordance with what we observed in the HD chromosomes, these unstable transmissions happened in normal chromosomes with haplotypes B and C, showing that in the non-pathological allele range instability also occurs in other haplotypes. Interestingly, in a total of 22 normal
chromosomes associated with the delta2642 deletion allele (that marks haplotype A and also the rare haplotypes E and G) all of the normal HTT CAG repeats were stably transmitted.

In our study, we observed high haplotype diversity in both normal and expanded chromosomes, representative of more than one ancestral chromosome underlying HD in Portugal, where multiple independent events have given rise to the expansion into HD. Normal HD alleles with higher CAG repeats had a higher similarity in the extended HTT 4p16.3 region with HD alleles, independent of the haplotype, suggesting as expected that these range of alleles are a reservoir for new HD mutations. Haplotype A, marked by the HTT delta2642 deletion, was enriched in HD chromosomes (4-fold increase) and seemed to be particularly prone to expansion, supported by the fact that control chromosomes with this haplotype had the highest median CAG repeat length and was associated with the two biggest expansions observed in the Portuguese population.

**Acknowledgments** The authors would like to thank all patients and their families who accepted to participate in this study. We also thank to all the physicians who referred HD testing to our center: A Ramos, A Fernandes, JM Leite, F Nogueira, S Beck, Mestre T, FR Martins, F Magalhães, GC Santo, B Botelho, G Neves, JC Pereira, E Monteiro, AS Morgadinho, AI Rebelo, AM Rodrigues, M Rodrigues, A Geraldo, B Rodrigues, C Januário, MA Ferro, MC Macário, NMM da Silva, MHC Fernandes, A Morganho, L Nunes, T Lourenço, J Guimarães, J Vale, MJ Feijó, P Tavares, MTC Pereira, AB Sousa, F Falcão, H Santos, I Cordeiro, IM Gaspar, J Ferreira, MM Rosa, M Lauterbach, R Geraldes, C Moura, C Garrett, JM Campillo, P Esperança, J Figueiredo, M Veloso, JL Loureiro, P Coutinho, F Piloto, A Henriques, AP Carvalho, C Moço, A Palos, J Becho, JMC Guerra, MG Mira, C Costa, J Coimbra, J Proença, JCP Ferreira, RJ Guerra, AP Correia, A Tuna, M Magalhães, M Mota, V Passão, FV Ramos, JJS de Sá, JM Saraiva, M Amorim, SB de Sousa, S Maia, D Alves, AC Araújo, H Águas, MS Dias, M Lourenço, M Rocha, D Martins and J Marinho.

**Funding** Financiamento Plurianual de Unidades de Investigação (FCT) and NIH grants from the NINDS NS16367 (The Massachusetts HD Center Without Walls) and NS32765. EMR is the recipient of a scholarship from Fundação para a Ciência e a Tecnologia (SFRH/BD/44335/2008). IA is funded by Programa Ciência, POPH – QREN – Tipologia 4.2 - Promoção do Emprego Científico, co-funded by ESF and national funds by the MCTES (Ministério da Ciência e Ensino Superior).
Disclosure of Interests None.

References


Genetic modifiers of Huntington’s disease


### Tables and Figures

#### Table 1. Transmission of the expanded HD repeat in major core haplotypes A to D

<table>
<thead>
<tr>
<th>Core haplotype</th>
<th>All</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maternal transmissions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transmissions, n</td>
<td>41</td>
<td>10</td>
<td>12</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Range of transmitted allele (CAGs)</td>
<td>40-73</td>
<td>42-73</td>
<td>40-54</td>
<td>40-65</td>
<td>41-46</td>
</tr>
<tr>
<td>(mean)</td>
<td>(44)</td>
<td>(47)</td>
<td>(43)</td>
<td>(45)</td>
<td>(43)</td>
</tr>
<tr>
<td>Unstable transmissions, n</td>
<td>28</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Contractions, n</td>
<td>15</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Expansions, n</td>
<td>13</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Repeat-length variation, mean</td>
<td>+0.59</td>
<td>+2.20</td>
<td>+0.50</td>
<td>-0.10</td>
<td>-0.13</td>
</tr>
<tr>
<td>Range</td>
<td>-2 to +25</td>
<td>-2 to +25</td>
<td>-2 to +4</td>
<td>-2 to +2</td>
<td>-2 to +1</td>
</tr>
<tr>
<td><strong>Paternal transmissions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transmissions, n</td>
<td>42</td>
<td>9</td>
<td>19</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Range of transmitted allele (CAGs)</td>
<td>36-52</td>
<td>41-52</td>
<td>36-51</td>
<td>39-46</td>
<td>38-43</td>
</tr>
<tr>
<td>(mean)</td>
<td>(42)</td>
<td>(44)</td>
<td>(42)</td>
<td>(42)</td>
<td>(42)</td>
</tr>
<tr>
<td>Unstable transmissions, n</td>
<td>31</td>
<td>8</td>
<td>12</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Contractions, n</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Expansions, n</td>
<td>27</td>
<td>8</td>
<td>9</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Repeat-length variation, mean</td>
<td>+2.69</td>
<td>+4.89</td>
<td>+2.00</td>
<td>+1.40</td>
<td>+1.71</td>
</tr>
<tr>
<td>Range</td>
<td>-2 to +24</td>
<td>0 to +24</td>
<td>-2 to +12</td>
<td>0 to +3</td>
<td>-1 to +7</td>
</tr>
</tbody>
</table>
Figure 1. Schematic map of the polymorphisms around the HD gene used for the haplotype reconstruction. A region of 1.17Mb flanking the HD gene is represented, showing the location of nine SNPs highly associated with HD chromosomes (rs6814736, rs7667745, rs1751848, rs762847, rs7658462, rs3129317, rs1730768, rs12641989 and rs16844364) and microsatellite marker (D4S127). In the HD gene, besides the CAG repeat itself it was also included in this study the adjacent proline repeat and the in/del polymorphism (delta2642). The four most common HD extended haplotypes (1 to 4) and the eight different core haplotypes (A to H) found on HD chromosomes are shown, along with the allele of each of the markers that define each haplotype.
Figure 2. Frequency of core haplotypes among HD and control chromosomes. The frequencies of each core haplotype (A to H) on HD (n=317) and control (n=469) chromosomes are shown, as well as the frequency of control chromosomes bearing “other” core haplotypes (light grey). The remaining (unfilled) represents the frequency of chromosomes where the core haplotype could not be unambiguously determined.
Figure 3. Overview of the extended ~1Mb region in each core haplotype (A to D) for both HD and control chromosomes. The darkest color corresponds to the genotypes that defined the core haplotype (around the HD CAG repeat) and are delimited by boxes. The light color represents the loci that shared the same allele as the correspondent extended haplotype (haplotype A: extended haplotype 1, haplotype B: extended haplotype 2, haplotype C: extended haplotype 4 and haplotype D: extended haplotype 3, as well as the rare haplotypes E and G: extended haplotype 1), while the grey color represents loci that have a different allele at that position. The white color corresponds to missing genotypes or not ambiguously phased loci. The chromosomes for each core haplotype were sorted based on the associated D4S127 allele (first chromosomes with: 157bp for haplotypes A, E and G; 155bp for haplotype B; 151bp for haplotypes C and D – represented in dark color; and for haplotype C followed by chromosomes with 155bp – in darker blue; then by chromosomes with other D4S127 genotypes – in grey and chromosomes with missing genotypes – in white) and then by similarity of the extended haplotype. Arrows mark normal alleles with more than 20 CAGs.
In our set of chromosomes, the median CAG repeat for core haplotype A was 42.0 (n=101), B was 24.5 (n=268), C was 19.5 (n=126) and D was 18.0 (n=155). And we observed the same trend, even by including only normal range allele sizes (<36 CAGs), were the median normal CAG size for core haplotype A was 26.0 (n=23), 18.0 for B (n=149), 17.0 for C (n=77) and 17.0 for D (n=119). For haplotype C, we separated the 126 chromosomes based on their D4S127 genotype: from bottom to top, 151bp in dark blue, 155bp in darker blue, other in grey and missing genotype in white. The median CAG repeat for haplotype C associated with 151bp D4S127 allele was 17.0 (n=70) and with 155bp was 42.5 (n=20).

**Figure 4.** HD CAG distribution for all chromosomes with the major core haplotypes A to D. In our set of chromosomes, the median CAG repeat for core haplotype A was 42.0 (n=101), B was 24.5 (n=268), C was 19.5 (n=126) and D was 18.0 (n=155). And we observed the same trend, even by including only normal range allele sizes (<36 CAGs), were the median normal CAG size for core haplotype A was 26.0 (n=23), 18.0 for B (n=149), 17.0 for C (n=77) and 17.0 for D (n=119). For haplotype C, we separated the 126 chromosomes based on their D4S127 genotype: from bottom to top, 151bp in dark blue, 155bp in darker blue, other in grey and missing genotype in white. The median CAG repeat for haplotype C associated with 151bp D4S127 allele was 17.0 (n=70) and with 155bp was 42.5 (n=20).
Figure 5. Pedigrees and haplotypes from two families in the MICROS study that carried unstable non-pathological HD alleles. Grey and colored symbols indicate samples that were genotyped, with the colored ones indicating individuals that carried an unstable allele (blue indicates that the unstable allele was associated with haplotype C and brown corresponds to haplotype B). Haplotypes of 4 markers spanning the HD gene (core-haplotype: rs762847 - D4S127 - HD CAG repeat - CCG repeat - delta2642, from top to bottom) are shown, with the unstable alleles haplotype boxed in the corresponding haplotype color (brown: haplotype B, blue: haplotype C).
### Supplementary Table 1. Percentage of the extended haplotypes in 193 HD chromosomes.

<table>
<thead>
<tr>
<th>Haplotype ID</th>
<th>Extended haplotype sequence</th>
<th>Number of chromosomes</th>
<th>Percentage</th>
<th>Core haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CGGA7delTGGAA</td>
<td>36</td>
<td>18.65</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>GCAA7nCGAGG</td>
<td>28</td>
<td>14.51</td>
<td>C</td>
</tr>
<tr>
<td>3</td>
<td>GCGG10nCAGGG</td>
<td>13</td>
<td>6.74</td>
<td>E</td>
</tr>
<tr>
<td>4</td>
<td>GCAG7nCGAGG</td>
<td>12</td>
<td>6.22</td>
<td>D</td>
</tr>
<tr>
<td>5</td>
<td>GCAA7nCGAGA</td>
<td>8</td>
<td>4.15</td>
<td>C</td>
</tr>
<tr>
<td>6</td>
<td>GCAA7nCAGGG</td>
<td>7</td>
<td>3.63</td>
<td>C</td>
</tr>
<tr>
<td>7</td>
<td>GCAA7nCAGGG</td>
<td>7</td>
<td>3.63</td>
<td>C</td>
</tr>
<tr>
<td>8</td>
<td>GGGAA7delTGGAA</td>
<td>7</td>
<td>3.63</td>
<td>A</td>
</tr>
<tr>
<td>9</td>
<td>GCA7G10nCAGGG</td>
<td>6</td>
<td>3.11</td>
<td>E</td>
</tr>
<tr>
<td>10</td>
<td>GGGG7nCAGGA</td>
<td>6</td>
<td>3.11</td>
<td>D</td>
</tr>
<tr>
<td>11</td>
<td>CGGA7delTGGGG</td>
<td>4</td>
<td>2.07</td>
<td>A</td>
</tr>
<tr>
<td>12</td>
<td>CGGA7nCGAGG</td>
<td>4</td>
<td>2.07</td>
<td>C</td>
</tr>
<tr>
<td>13</td>
<td>GCGG7nCGGGG</td>
<td>4</td>
<td>2.07</td>
<td>C</td>
</tr>
<tr>
<td>14</td>
<td>GCGG7nCAGGGG</td>
<td>4</td>
<td>2.07</td>
<td>D</td>
</tr>
<tr>
<td>15</td>
<td>GCGG7nTGGAGG</td>
<td>4</td>
<td>2.07</td>
<td>D</td>
</tr>
<tr>
<td>16</td>
<td>GGA7nCGAGG</td>
<td>4</td>
<td>2.07</td>
<td>C</td>
</tr>
<tr>
<td>17</td>
<td>CCGG7nCAGGG</td>
<td>3</td>
<td>1.55</td>
<td>D</td>
</tr>
<tr>
<td>18</td>
<td>CGGA7delTGGAG</td>
<td>3</td>
<td>1.55</td>
<td>A</td>
</tr>
<tr>
<td>19</td>
<td>GCA7G10nCAGGA</td>
<td>3</td>
<td>1.55</td>
<td>C</td>
</tr>
<tr>
<td>20</td>
<td>GGA7nCAGAGA</td>
<td>3</td>
<td>1.55</td>
<td>C</td>
</tr>
<tr>
<td>21</td>
<td>GGA7nCAGGA</td>
<td>3</td>
<td>1.55</td>
<td>C</td>
</tr>
<tr>
<td>22</td>
<td>CGGA7delTGGAA</td>
<td>2</td>
<td>1.04</td>
<td>A</td>
</tr>
<tr>
<td>23</td>
<td>GCA7G8nCAGGG</td>
<td>2</td>
<td>1.04</td>
<td>F</td>
</tr>
<tr>
<td>24</td>
<td>GCAG7delCAGGG</td>
<td>2</td>
<td>1.04</td>
<td>A</td>
</tr>
<tr>
<td>25</td>
<td>GCA7G7nCAGGG</td>
<td>2</td>
<td>1.04</td>
<td>C</td>
</tr>
<tr>
<td>26</td>
<td>GCGG7nTGGAGG</td>
<td>2</td>
<td>1.04</td>
<td>D</td>
</tr>
<tr>
<td>27</td>
<td>GGA7nCAGGA</td>
<td>2</td>
<td>1.04</td>
<td>C</td>
</tr>
<tr>
<td>28</td>
<td>CGGA7delCAGGG</td>
<td>1</td>
<td>0.52</td>
<td>A</td>
</tr>
<tr>
<td>29</td>
<td>GCA7G7nCAGGG</td>
<td>1</td>
<td>0.52</td>
<td>C</td>
</tr>
<tr>
<td>30</td>
<td>GCA7G7delTGGAA</td>
<td>1</td>
<td>0.52</td>
<td>A</td>
</tr>
<tr>
<td>31</td>
<td>GCA7G10delCAGGG</td>
<td>1</td>
<td>0.52</td>
<td>G</td>
</tr>
<tr>
<td>32</td>
<td>GCAG7delTGGAA</td>
<td>1</td>
<td>0.52</td>
<td>B</td>
</tr>
<tr>
<td>33</td>
<td>GCAG7delTGGAA</td>
<td>1</td>
<td>0.52</td>
<td>A</td>
</tr>
<tr>
<td>34</td>
<td>GCGG7nCAGGA</td>
<td>1</td>
<td>0.52</td>
<td>C</td>
</tr>
<tr>
<td>35</td>
<td>GCGG7nCAGGA</td>
<td>1</td>
<td>0.52</td>
<td>D</td>
</tr>
<tr>
<td>36</td>
<td>GCGG7nCAGGG</td>
<td>1</td>
<td>0.52</td>
<td>D</td>
</tr>
<tr>
<td>37</td>
<td>GGA7G7delTGGAA</td>
<td>1</td>
<td>0.52</td>
<td>A</td>
</tr>
<tr>
<td>38</td>
<td>GGA7nCAGGG</td>
<td>1</td>
<td>0.52</td>
<td>C</td>
</tr>
<tr>
<td>39</td>
<td>GGGAA7delTGGAG</td>
<td>1</td>
<td>0.52</td>
<td>A</td>
</tr>
</tbody>
</table>
Supplementary Table 2. Percentage of the extended haplotypes in 267 control chromosomes.

<table>
<thead>
<tr>
<th>Haplotype ID</th>
<th>Extended haplotype sequence</th>
<th>Number of chromosomes</th>
<th>Percentage</th>
<th>Core haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>GCAG10nCAGGG</td>
<td>41</td>
<td>15.36</td>
<td>E</td>
</tr>
<tr>
<td>20</td>
<td>GCAG7nCAGGG</td>
<td>21</td>
<td>7.87</td>
<td>C</td>
</tr>
<tr>
<td>2</td>
<td>GCAA7nCGAGG</td>
<td>19</td>
<td>7.12</td>
<td>C</td>
</tr>
<tr>
<td>3</td>
<td>GCGG10nCAGGG</td>
<td>15</td>
<td>5.62</td>
<td>E</td>
</tr>
<tr>
<td>14</td>
<td>GCGG7nCAGGG</td>
<td>14</td>
<td>5.24</td>
<td>D</td>
</tr>
<tr>
<td>6</td>
<td>GCAA7nCAGGG</td>
<td>13</td>
<td>4.87</td>
<td>C</td>
</tr>
<tr>
<td>1</td>
<td>CGGA7delTGGAA</td>
<td>10</td>
<td>3.75</td>
<td>A</td>
</tr>
<tr>
<td>36</td>
<td>GCGG7nCGAGG</td>
<td>8</td>
<td>3.00</td>
<td>D</td>
</tr>
<tr>
<td>8</td>
<td>GCGA7delTGGAA</td>
<td>7</td>
<td>2.62</td>
<td>A</td>
</tr>
<tr>
<td>40</td>
<td>GCAA8nCGAGA</td>
<td>7</td>
<td>2.62</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>GCAA7nCGGGG</td>
<td>6</td>
<td>2.25</td>
<td>C</td>
</tr>
<tr>
<td>41</td>
<td>GCAG10nTAGGG</td>
<td>6</td>
<td>2.25</td>
<td>E</td>
</tr>
<tr>
<td>4</td>
<td>GCAG7nCAGGG</td>
<td>5</td>
<td>1.87</td>
<td>D</td>
</tr>
<tr>
<td>13</td>
<td>GCAA7nCGGGG</td>
<td>4</td>
<td>1.50</td>
<td>C</td>
</tr>
<tr>
<td>42</td>
<td>GCAG7nTAGGG</td>
<td>4</td>
<td>1.50</td>
<td>D</td>
</tr>
<tr>
<td>43</td>
<td>GCGG10nCGAGG</td>
<td>4</td>
<td>1.50</td>
<td>E</td>
</tr>
<tr>
<td>44</td>
<td>GGA7nCAGGG</td>
<td>4</td>
<td>1.50</td>
<td>C</td>
</tr>
<tr>
<td>45</td>
<td>GCGG10nCAGGG</td>
<td>4</td>
<td>1.50</td>
<td>E</td>
</tr>
<tr>
<td>23</td>
<td>GCAG8nCAGGG</td>
<td>3</td>
<td>1.12</td>
<td>F</td>
</tr>
<tr>
<td>33</td>
<td>GCGA7delTGGAA</td>
<td>3</td>
<td>1.12</td>
<td>A</td>
</tr>
<tr>
<td>46</td>
<td>GCAG10nCAGGG</td>
<td>3</td>
<td>1.12</td>
<td>E</td>
</tr>
<tr>
<td>47</td>
<td>GCGA10nCAGGA</td>
<td>3</td>
<td>1.12</td>
<td>H</td>
</tr>
<tr>
<td>48</td>
<td>GCGA7nCAGGA</td>
<td>3</td>
<td>1.12</td>
<td>C</td>
</tr>
<tr>
<td>49</td>
<td>GCGA8nCAGGG</td>
<td>3</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>GCAA7nCAAGG</td>
<td>3</td>
<td>1.12</td>
<td>C</td>
</tr>
<tr>
<td>34</td>
<td>GCAG7nCGAGA</td>
<td>2</td>
<td>0.75</td>
<td>C</td>
</tr>
<tr>
<td>51</td>
<td>CCGG7nCAGAG</td>
<td>2</td>
<td>0.75</td>
<td>D</td>
</tr>
<tr>
<td>52</td>
<td>GAA7nCAGAG</td>
<td>2</td>
<td>0.75</td>
<td>C</td>
</tr>
<tr>
<td>53</td>
<td>CGAG7nTGGGA</td>
<td>2</td>
<td>0.75</td>
<td>D</td>
</tr>
<tr>
<td>54</td>
<td>GCAA10nCAGGG</td>
<td>2</td>
<td>0.75</td>
<td>H</td>
</tr>
<tr>
<td>55</td>
<td>GCAG10nCAGAA</td>
<td>2</td>
<td>0.75</td>
<td>E</td>
</tr>
<tr>
<td>56</td>
<td>GCAG10nCAGGA</td>
<td>2</td>
<td>0.75</td>
<td>E</td>
</tr>
<tr>
<td>57</td>
<td>GCAG8nTGGGA</td>
<td>2</td>
<td>0.75</td>
<td>F</td>
</tr>
<tr>
<td>58</td>
<td>GCAG8nTGGGG</td>
<td>2</td>
<td>0.75</td>
<td>F</td>
</tr>
<tr>
<td>59</td>
<td>GCGA10nCAGGG</td>
<td>2</td>
<td>0.75</td>
<td>H</td>
</tr>
<tr>
<td>60</td>
<td>GCAG7nTGGGG</td>
<td>2</td>
<td>0.75</td>
<td>C</td>
</tr>
<tr>
<td>12</td>
<td>CCGA7nCGAGG</td>
<td>1</td>
<td>0.37</td>
<td>C</td>
</tr>
<tr>
<td>16</td>
<td>GAA7nCAGAG</td>
<td>1</td>
<td>0.37</td>
<td>C</td>
</tr>
<tr>
<td>29</td>
<td>GCGA7nCAGGG</td>
<td>1</td>
<td>0.37</td>
<td>C</td>
</tr>
<tr>
<td>61</td>
<td>CCGA7nCGAGG</td>
<td>1</td>
<td>0.37</td>
<td>C</td>
</tr>
<tr>
<td>62</td>
<td>CCGA7delTGGAA</td>
<td>1</td>
<td>0.37</td>
<td>A</td>
</tr>
<tr>
<td>63</td>
<td>CCGA7nCAGGA</td>
<td>1</td>
<td>0.37</td>
<td>C</td>
</tr>
<tr>
<td>64</td>
<td>GCAG10nCAGGG</td>
<td>1</td>
<td>0.37</td>
<td>E</td>
</tr>
<tr>
<td>65</td>
<td>GCGA7nCAGGG</td>
<td>1</td>
<td>0.37</td>
<td>C</td>
</tr>
<tr>
<td>Haplotype ID</td>
<td>Extended haplotype sequence</td>
<td>Number of chromosomes</td>
<td>Percentage</td>
<td>Core haplotype</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------</td>
<td>-----------------------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td>66</td>
<td>CGGA7nTGGGG</td>
<td>1</td>
<td>0.37</td>
<td>C</td>
</tr>
<tr>
<td>67</td>
<td>CGGG10nCAGGG</td>
<td>1</td>
<td>0.37</td>
<td>E</td>
</tr>
<tr>
<td>68</td>
<td>CGGG7nCGAGA</td>
<td>1</td>
<td>0.37</td>
<td>D</td>
</tr>
<tr>
<td>69</td>
<td>GCAA7delTGAGG</td>
<td>1</td>
<td>0.37</td>
<td>A</td>
</tr>
<tr>
<td>70</td>
<td>GCAAnCAAGA</td>
<td>1</td>
<td>0.37</td>
<td>C</td>
</tr>
<tr>
<td>71</td>
<td>GCAA7nCAAGG</td>
<td>1</td>
<td>0.37</td>
<td>C</td>
</tr>
<tr>
<td>72</td>
<td>GCAA7nTAGGG</td>
<td>1</td>
<td>0.37</td>
<td>C</td>
</tr>
<tr>
<td>73</td>
<td>GCAA9nCAAGG</td>
<td>1</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>GCAA9nCAGGG</td>
<td>1</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>GCAAnCGGG</td>
<td>1</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>GCAG10nCGAGA</td>
<td>1</td>
<td>0.37</td>
<td>E</td>
</tr>
<tr>
<td>77</td>
<td>GCAG7nCAAGG</td>
<td>1</td>
<td>0.37</td>
<td>D</td>
</tr>
<tr>
<td>78</td>
<td>GCAG7nCAGGG</td>
<td>1</td>
<td>0.37</td>
<td>D</td>
</tr>
<tr>
<td>79</td>
<td>GCAG7nCGAGA</td>
<td>1</td>
<td>0.37</td>
<td>D</td>
</tr>
<tr>
<td>80</td>
<td>GCAG7nTGAGA</td>
<td>1</td>
<td>0.37</td>
<td>D</td>
</tr>
<tr>
<td>81</td>
<td>GCAG7nTGGGG</td>
<td>1</td>
<td>0.37</td>
<td>D</td>
</tr>
<tr>
<td>82</td>
<td>GCAG7nCAAGG</td>
<td>1</td>
<td>0.37</td>
<td>C</td>
</tr>
<tr>
<td>83</td>
<td>GCAAnCAGAA</td>
<td>1</td>
<td>0.37</td>
<td>C</td>
</tr>
<tr>
<td>84</td>
<td>GCAAnCAAGG</td>
<td>1</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>GCGG10nCAGGA</td>
<td>1</td>
<td>0.37</td>
<td>E</td>
</tr>
<tr>
<td>86</td>
<td>GCGG7nTAGGA</td>
<td>1</td>
<td>0.37</td>
<td>D</td>
</tr>
<tr>
<td>87</td>
<td>GCGG7nCAGGG</td>
<td>1</td>
<td>0.37</td>
<td>D</td>
</tr>
<tr>
<td>88</td>
<td>GGGG10nCAGGG</td>
<td>1</td>
<td>0.37</td>
<td>E</td>
</tr>
<tr>
<td>89</td>
<td>GGGG7nCAGGG</td>
<td>1</td>
<td>0.37</td>
<td>D</td>
</tr>
</tbody>
</table>
Huntington’s Disease *trans*-modifiers

**Huntington's Disease candidate modifier genes**

In the search for modifiers, about 20 *loci* have been implicated in AO or progression of HD, though with different results in different studies. We have critically evaluated, in one of the largest collection of HD samples of European origin, the effect of some of the most compelling candidates on residual AO (corrected for CAG size) of HD symptoms. Additionally, we assessed the potential role of variants in genes of the dopamine pathway known to have a functional impact in a pathway that is known to be disrupted in HD, and also the contribution of expanded polyQ repeats known to cause other neurodegenerative disorders.

Our thorough analysis of the effect of rs7665116 in the mitochondrial regulator *PPARGC1A* gene, revealed that when examined by origin, our Southern and Western European HD samples differed for minor allele frequency (MAF) - consistent with an ancestry tagging-SNP; and, despite similar mean CAG sizes, Southern European samples had an older mean AO - revealing phenotype stratification [Article 5]. Therefore and in order to avoid the dramatic effect of phenotypic (AO) and genotypic (MAF) stratification in our HD association studies, we used a population-specific regression analysis, where models were adjusted for ancestry. These candidate genes association studies did not provide strong evidence to support a modulator role in AO of HD motor manifestations for variations neither within the *PPARGC1A* [Article 5] nor the glutamatergic and dopaminergic genes [Article 6]. We also did not find an evident genetic association of HD with expanded CAG repeats in any of the polyQ disease genes analyzed. However, a small proportion of the HD subjects carried extremely rare *SCA2* intermediate alleles, recently shown to be associated with other neurological diseases [Article 7].
Article 5: Population stratification may bias analysis of PGC-1α as a modifier of age at Huntington disease motor onset.


Reprinted with the permission of Springer
License number 3190400940558
Population stratification may bias analysis of PGC-1α as a modifier of age at Huntington disease motor onset

Eliana Marisa Ramos · Jeanne C. Latourelle · Ji-Hyun Lee · Tammy Gillis · Jyvalakshmi S. Mysore · Ferdinando Squitieri · Alba Di Pardo · Stefano Di Donato · Michael R. Hayden · Patrick J. Morrison · Martha Nance · Christopher A. Ross · Russell L. Margolis · Estrella Gomez-Tortosa · Carmen Ayuso · Oksana Suchowersky · Ronald J. Trent · Elizabeth McCusker · Andrea Novello · Marina Frontali · Randi Jones · Tetsuo Ashizawa · Samuel Frank · Mario-Helene Saint-Hilaire · Steven M. Hersch · Hermínia D. Rosas · Diane Lucente · Madaline B. Harrison · Andrea Zanko · Karen Marder · James F. Gusella · Jong-Min Lee · Isabel Alonso · Jorge Sequeiros · Richard H. Myers · Marcy E. MacDonald

Received: 23 March 2012 / Accepted: 11 July 2012 / Published online: 25 July 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract Huntington’s disease (HD) is an inherited neurodegenerative disorder characterized by motor, cognitive and behavioral disturbances, caused by the expansion of a CAG trinucleotide repeat in the HD gene. The CAG allele size is the major determinant of age at onset (AO) of motor symptoms, although the remaining variance in AO is highly heritable. The rs7665116 SNP in PPARGC1A, encoding the mitochondrial regulator PGC-1α, has been reported to be a significant modifier of AO in three European HD cohorts, perhaps due to affected cases from Italy. We attempted to replicate these findings in a large collection of (1,727) HD patient DNA samples of European origin. In the entire
cohort, rs7665116 showed a significant effect in the dominant model (p value = 0.008) and the additive model (p value = 0.009). However, when examined by origin, cases of Southern European origin had an increased rs7665116 minor allele frequency (MAF), consistent with this being an ancestry-tagging SNP. The Southern European cases, despite similar mean CAG allele size, had a significantly older mean AO (p < 0.001), suggesting population-dependent phenotype stratification. When the generalized estimating equations models were adjusted for ancestry, the effect of the rs7665116 genotype on AO decreased dramatically. Our results do not support rs7665116 as a modifier of AO of motor symptoms, as we found evidence for a dramatic effect of phenotypic (AO) and genotypic (MAF) stratification among European cohorts that was not considered in previously reported association studies. A significantly older AO in Southern Europe may reflect population differences in genetic or environmental factors that warrant further investigation.

Introduction

Huntington’s disease (HD) is a neurodegenerative disorder with classic symptoms that include progressive chorea, motor deficits, cognitive changes and dementia. Age at onset (AO) of the overt symptoms is highly variable: while some individuals show signs in the first decade, others remain asymptomatic even after 60 years of age, though in most cases death follows on average about 17 years after the first symptoms. HD is inherited as an autosomal dominant trait and is caused by the expansion of an unstable CAG repeat, in the first exon of the HD gene (now called HTT), on chromosome 4p16.3 (The Huntington’s Disease Collaborative Research Group 1993), resulting in an expanded polyglutamine tract in the huntingtin protein.

The major determinant of AO in HD is the size of the expanded CAG repeat allele (Lee et al. 2012), such that the longer the repeat the earlier the onset of clinical symptoms, though most HD cases occur in adulthood with about 40–45 CAG repeats. Repeat length alone explains about 70% of the variability in onset age (Duyao et al. 1993). The remaining variance in AO (residual AO) is highly heritable but remains unexplained. Nonetheless, recent genetic studies have nominated about 20 loci that may modify AO or progression of HD. However, many of the specific polymorphisms assessed in multiple studies have failed to be replicated, including attractive biological candidate genes such as glutamate receptor, ionotropic kainate 2 (GRIK2), apolipoprotein E (APOE) and brain-derived neurotrophic factor (BDNF) (Rubinsztain et al. 1997; Metzger et al. 2006; Panas et al. 1999; Alberch et al. 2005; Di Maria et al. 2006).

One biologically compelling candidate thought to be involved in HD pathogenesis is PPARGC1A, localized at 4p15.1-2, which encodes peroxisome proliferator-activated receptor γ coactivator 1α (PGC-1α), a transcriptional regulator of adaptive thermogenesis (Puigserver et al. 1998) and mitochondrial respiration and oxidative stress (Puigserver and Spiegelman 2003; St-Pierre et al. 2006). The lack of PGC-1α expression produces an HD-like phenotype in mice (Lin et al. 2004; Leone et al. 2005) and mutant, but
Genetic modifiers of Huntington’s disease

Results – Article 5

not wild type, huntingtin down-regulates the expression of PGC-1α and its target genes (Cui et al. 2006; Lin et al. 2004; Weydt et al. 2006). Moreover, three recent studies, with DNA samples from European HD patients, mainly from Italy and Germany, have reported an association of PGC-1α with AO of HD symptoms (Che et al. 2011; Weydt et al. 2009; Taherzadeh-Fard et al. 2009). Although different sets of PPARC1A SNPs were included in these studies, one polymorphism (rs7665116), located at the 3′-end region of intron 2, was associated with a later AO in all three studies, displaying a significant effect both in an additive and dominant model. Following these results, it was reported that polymorphisms in PGC-1α downstream target genes, namely nuclear respiratory factor 1 (NRF1) and mitochondrial transcription factor A (TFAM), may influence the AO in HD (Taherzadeh-Fard et al. 2011). Given that results based on analyses of the PPARC1A rs7665116 SNP are motivating a broader range of research into the functional basis of the effect, the aim of the present study was to attempt to replicate the association of this SNP with AO, in a much larger cohort of 1,727 HD patients of different European populations.

Methods

Subjects

We analyzed 1,929 HD patients with known AO of overt motor symptoms. The DNA samples were from subjects involved in long-term genetic studies from collaborating investigators (HD-MAPS), the HD observational study COHORT and from the Harvard Tissue Resource Center Bank (McLean’s Hospital, Belmont MA) and the National Neurological Research Bank (VAMC Wadsworth Division, Los Angeles CA). These studies included related individuals [from 1,676 different families defined either based on the likelihood of genetic similarity from genome-wide genotyping information (Western European samples) or membership in nuclear (parents and children) families (Southern European samples)]. Of these, 934 were self-reported as originally from Southern European countries (263 from Portugal, 664 from Italy, 5 from Spain and 2 from Greece), the rest of the cases had unconfirmed or no geographical origin data. 1,020 of these were genotyped using the GeneChip Human Mapping 500K Array Set (Affymetrix) at the Broad Institute of Harvard and MIT as part of a genome-wide scan for HD genetic modifiers.

Genotyping

The HD CAG repeat length was determined by a polymerase chain reaction (PCR) amplification assay, using fluorescently labeled primers, as previously described (Warner et al. 1993). The size of the fragments was determined using the ABI PRISM 3730x/ automated DNA Sequencer (Applied Biosystems, Foster City, CA, USA) and GeneMarker version 3.7 software. A set of HD CAG alleles, determined by DNA sequencing, were used as standards. Genotyping of the PGC-1α polymorphism (rs7665116) was performed by real-time PCR using the commercially available Taqman Genotyping probe (Applied Biosystems, Foster City, CA, USA) carried out on the LightCycler480 (Roche Diagnostics, Mannheim), following manufacturer’s instructions.

Statistics

For the 1,020 samples with whole-genome genotyping, PCA was carried out using PLINK v1.05 (http://pngu.mgh.harvard.edu/Purcell/plink/) (Purcell et al. 2007) in order to determine the genetic ancestry of these individuals. Briefly, genotypes of HD samples were combined with HapMap Phase 2 data (CEPH, Yoruba, Han-Chinese and Japanese populations) for pairwise IBD estimation and subsequent IBS clustering.

To assess differences in the mean motor AO among Western and Southern European samples, we used the general estimating equation (GEE), thereby adjusting for related samples. Multivariate analyses were generated using GEE to assess the effect of the rs7665116 SNP at the PGC-1α gene with HD residual motor onset, adjusting for familial correlation. Residual motor onsets were computed as the difference between the observed and expected age of onset and were standardized to a mean of zero and standard deviation of one. The weighted GEE was computed assuming an independent correlation structure and using the robust estimator of the variance to account for familial relationships. All statistical analyses were performed using PASW Statistics (version 18).

Results

We genotyped a collection of 1,929 HD DNA samples, with known HD CAG allele sizes and known age at onset of motor symptoms, for the PPARC1A rs7665116 polymorphism. The observed genotype frequency of this SNP was in Hardy–Weinberg equilibrium. Since, in two of the previous reports, the association with AO was primarily observed in HD patients of Italian ancestry (Che et al. 2011; Weydt et al. 2009), we split our large cohort by ancestry into either Southern European or Western European HD cases. The Southern European HD cases (n = 934) consisted of self-reported Portuguese (n = 263), Italian (n = 664), Spanish (n = 5) and Greek (n = 2) HD cases. The
Western European HD cases were chosen from amongst another 1,020 HD patients by use of principal component analysis (PCA) on available whole-genome genotyping data, to infer their genetic background. The first principal component (PC1) distinguished Africans from non-Africans and the second principal component (PC2) distinguished Africans and Europeans from Asians (data not shown), and allowed us to exclude from our analysis the few samples who had significant contribution of either Asian or African ancestry. Among the remaining (n = 952) European cases, the Western European cluster (n = 793) was defined by overlap with the US Northern-Western European origin CEPH (HapMap) cluster, and consisted mainly of persons with self-reported North-American origin (Canada and US) as well as French and Irish. Thus, we had a total of 1,727 HD patients with assigned ancestry; 934 Southern European and 793 Western European (Table 1).

Remarkably, analysis of the clinical data for these 1,727 HD cases, which had CAG alleles ranging from 36 to 87 repeats, revealed that the self-reported Southern Europeans (n = 934) had significantly later onset of motor HD symptoms (p < 0.001), by 4–5 years, compared to the Western European (n = 793), though the means/medians for HD CAG repeat length were similar in both groups (Table 1). Furthermore, the observed rs7665116 genotypes for samples from the Southern European countries revealed higher minor allele frequency (MAF) (~17%) when compared to the Western European set (~12%) (Table 1). These findings, together with the striking differences in the MAF of this polymorphism among the different HapMap populations, strongly suggested that population stratification might increase type I errors in the AO association analysis.

Finally, we have recently shown that the non-normal distribution of CAG allele size (and AO) also introduces error in conventional statistical analysis (Lee et al. 2012). Even a single CAG outlier sample, with a very long CAG repeat and extremely young age at onset relative to all others, can have a profound effect on the final result when testing for the effects of potential genetic modifiers (Lee et al. 2012). Therefore, as a final filter, we chose only the Southern and Western European HD cases with CAG alleles in the 40- to 53-repeat range, shown previously to yield a statistically well-behaved data set that conforms to the fundamental assumptions of linear regression analysis (constant variance and normally distributed error) (Lee et al. 2012).

As summarized in Table 1, these filtering steps yielded a total of 879 self-reported Southern European cases from 823 families and 749 Western European cases (matched by use of PCA) from 620 different families, with CAG repeats ranging from 40 to 53 repeats. Notably, as observed for the larger set of HD patients with a broader CAG repeat range, in this final set of 1,628 patients the self-reported Southern Europeans had a significantly older mean age at onset of motor symptoms than Western Europeans (Table 1), which was observed across the spectrum of CAG allele sizes (Fig. 1), despite a similar mean/median HD CAG repeat

![Image](https://via.placeholder.com/150)

Fig. 1 Variance of age at motor onset for HD cases of Western and Southern European origin. A box plot depicting the relationship of the natural log-transformed age at onset of motor symptoms to the expanded CAG allele size, for patients in the 40–53 CAG range, illustrating that self-reported Southern Europeans had an older age at onset across the spectrum of allele sizes. Circles are outliers defined by a standard interquartile method (outside of 1.5 times interquartile range), some of which could reflect errors in the motor AO data while others may represent true biological outliers.
size (Table 1). Furthermore, the rs7665116 MAF was higher in the former relative to the latter Western European patients (Table 1).

Using this final set of 1,628 HD patients, we then performed analysis to determine whether rs7665116 may contribute to variance in HD motor onset not explained by the length of the expanded CAG repeat. In order to adjust for familial relationships, the effect of the rs7665116 on residual motor onset was calculated using generalized estimating equations (GEE). In the unadjusted analysis, a significant association with later residual AO was observed for both the additive genetic model ($\beta = 0.090$, $p$ value = 0.009) and the dominant model ($\beta = 0.113$, $p$ value = 0.008) (Table 2). However, adjusting the analysis for ancestry (Southern vs. Western European), in both the additive and the dominant models, produced a striking impact on the effect sizes ($\beta$ decreased by $\sim 25\%$) and the $p$ values ($p$ increased $\sim 4\times$) (Table 2), thereby revealing that population stratification is a large contributor to an apparent rs7665116 association.

**Discussion**

Previous studies have reported the presence of a common polymorphism in **PPARGC1A** (rs7665116) that is associated with a delay in AO of HD motor symptoms in three European HD cohorts (Che et al. 2011; Taherzadeh-Fard et al. 2009; Weydt et al. 2009), primarily contributed by patients from Italy. Our study, which involved a larger collection of HD cases, did not provide strong evidence for this SNP, and therefore, for **PGC-1a** as a modifier of HD motor onset, but did strongly support further investigation of the factors that contribute to the striking differences in AO of motor symptoms in ‘Southern Europeans’.

The results of our study expose genetic ancestry as a critical factor in HD association studies. It is expected that a disease-associated polymorphism may have varying effects in different populations, but the variation in minor allele frequency, across different genetic backgrounds related to ancestry may also be critical and should be taken into account in genetic association analysis. Cases that are poorly matched for genetic background may lead to false positives in association studies. It is important to control for ancestry by use of PCA or related methods, even in apparently close related populations such as Europeans (Novembre et al. 2008) where there is strong evidence of recent population selection that has led to intra-European variation in allele frequency (Price et al. 2008). This is particularly important in studies using the CEPH-CEU panel as controls since the genetic matching to different populations may differ considerably in different countries (Luo et al. 2008). It is difficult to accurately infer ancestry in candidate gene association studies, leading to imperfect correction for stratification. However, with increasing availability of genome-wide datasets, the assessment of population structure should become a common procedure for candidate association studies.

Our study points out the dramatic effect that population stratification can have in testing a candidate gene for an association with disease phenotype. We found that the variation in rs7665116 minor allele frequency could lead to a false positive, if genetic ancestry is not corrected for in the analysis. The Southern European cases seem to be different genetically and clinically, in terms of age at diagnosis of motor symptoms, from other European samples. By adjusting for ancestry, we observed striking effects on both the $p$ values (increased $\sim 4\times$) and effect sizes (decreased by $\sim 25\%$). Even though the post-adjustment $p$ values remained nominally significant, the dramatic reduction in significance occasioned by considering ancestry does not

### Table 2: Multivariate correlation of rs7665116 with residual age at motor onset

<table>
<thead>
<tr>
<th>Model</th>
<th>Standardized coefficient</th>
<th>Standard error</th>
<th>95% confidence limits</th>
<th>$p$ value*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Additive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PGC1α rs7665116</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T-T vs. T-C vs. C-C</td>
<td>0.090</td>
<td>0.0348</td>
<td>0.022</td>
<td>0.158</td>
</tr>
<tr>
<td><strong>PGC1α rs7665116 + ancestry</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T-T vs. T-C vs. C-C</td>
<td>0.069</td>
<td>0.0345</td>
<td>0.001</td>
<td>0.136</td>
</tr>
<tr>
<td>WE vs. SE</td>
<td>0.208</td>
<td>0.0416</td>
<td>0.127</td>
<td>0.290</td>
</tr>
<tr>
<td><strong>Dominant</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PGC1α rs7665116</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T-T vs. T-C vs. C-C</td>
<td>0.113</td>
<td>0.0424</td>
<td>0.029</td>
<td>0.196</td>
</tr>
<tr>
<td><strong>PGC1α rs7665116 + ancestry</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T-T vs. T-C vs. C-C</td>
<td>0.089</td>
<td>0.0418</td>
<td>0.007</td>
<td>0.171</td>
</tr>
<tr>
<td>WE vs. SE</td>
<td>0.208</td>
<td>0.0415</td>
<td>0.127</td>
<td>0.289</td>
</tr>
</tbody>
</table>

* $p$ values were derived using GEE to account for familial relationships.
lend confidence to its being a true effect and rather suggests that it is due to insufficiently rigorous ancestry categorization of the ‘Southern European’ set, which was based solely on self-reporting rather than unbiased genome-wide genotyping data. Consistent with this interpretation, while this manuscript was under review, an article by Soyal et al. (2012) reported that no association of rs7651166 with AO to first symptom (not necessarily motor) was found in an European cohort of 1,706 HD patients whose MAF for this SNP closely resembled the MAF that we report for the Western European samples in our study.

A remarkable finding from our analysis is the strong evidence of later motor onset for HD patients originally from Southern European countries (Portugal and Italy), which have a reported onset of motor symptoms that is 4–5 years later than that of HD patients from other European regions, despite similar mean/median CAG allele size. This striking difference may reflect population differences elsewhere in the genome, since extensive genome-wide SNP analyses have shown that even though European populations share much of their genetic background, they also exhibit a notable degree of non-sharing (ancestry) (Lao et al. 2008; Novembre et al. 2008). It’s important to note that in addition to genetic background differences, a number of other factors may contribute to the difference in age at onset. One possibility is environmental influences, for example, differences related to lifestyle, or perhaps types of medication. A meta-analysis study has found evidence that people who adhere to a Mediterranean Diet appear to have a reduced risk of developing Parkinson’s and Alzheimer’s disease (Soff et al. 2008), and altered Alzheimer’s disease course (Scarneceas et al. 2007). A recent observational study of HD in Europe has shown that Southern European (and Polish) clinicians prescribed anti-dyskinetic medication more frequently than clinicians in other European regions (Orth et al. 2010). Another environment related factor that is likely to contribute are the criteria and procedures for diagnosing HD, which may differ in different cultures. It is now important to understand which of many potential population-specific genetic and/or environmental factors are associated with later reported AO of motor symptoms in Southern Europeans.

**Conclusion**

The results of our study do not provide strong evidence for PPARGC1A SNP rs7665116, and therefore, for PGC-1α, as a modifier of age at onset of HD motor symptoms. However, we have found evidence of a significantly later age at onset of motor symptoms in Southern European countries, which may reflect genetic effects and/or environmental (lifestyle, diagnosis) factors that should be further explored.

Our data strongly illustrate the false contribution that population stratification may make in a candidate gene association study, while providing genetic evidence that the contribution of PGC-1α as a modifier of the disease process that leads to onset of HD motor symptoms may not be significant.

**Acknowledgments** We thank the HD families, whose participation in genetic studies made this work possible, the COHORT co-investigators and contributors (see “Appendix”), contributors to the HD-MAPS study, Ruth Abramson, Alexandra Durr, Adam Rosenblatt, Luigi Frati, Sasan Perlman, P. Michael Conneally, Mary Lou Klimmek, Melissa Diggins, Tiffany Hadzi and Ayana Dukett, as well as the Harvard Brain Tissue Resource Center at McLean Hospital and the National Neurological Research Specimen Bank at the VA West Los Angeles Health Care Center. EMR is the recipient of a scholarship from the Fundação para a Ciência e a Tecnologia of Portugal (SFRH/BID/44335/2008). Study funding supported by NIH grants from the NINDS NS16367 (The Massachusetts HD Center Without Walls) and NS32765, and the CHDI Foundation, Inc.

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical Standards** This study used only deidentified, previously collected DNA samples and phenotypic data in a manner approved by the Institutional Review Board of Partners HealthCare, Inc.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

**Appendix: Huntington Study Group COHORT Investigators**

**Steering Committee**

University of Rochester, Rochester, NY: Ira Shoulson, MD (principal investigator), James Gusella, PhD (co–principal investigator), Tatiana Foroud (co–principal investigator), Irina Antonijevic, Dan van Knammen (CHDI Foundation Inc.).

**Publications and Data Use Committee**

Tatiana Foroud (Chair), Ray Dorsey (Co-Chair), John Warner (CHDI), Joe Giuliano, Louise Vetter, Oksana Suchowskyy, Christopher Beck, David Oakes.

**Participating Investigators and Coordinators**

F. Marshall, University of Rochester; K. Marder, S. Frucht, C. Moskowitz, R. Clouse, P. Wasserman, Columbia University Medical Center; K. Shannon, J. Jaglin, Rush University Medical Center;
Results – Article 5

References


Genetic modifiers of Huntington’s disease


Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maler J, Sklar P, de Bakker PI, Daly MJ, Shum PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3):559–575. doi:10.1016/j.ajhg.2007.08.010


Springer
Article 6: Candidate glutamatergic and dopaminergic pathway gene variants do not influence age at motor onset of Huntington’s disease.


Neurogenetics [in press]

Reprinted with the permission of Springer
License number 3190400840043
Candidate glutamatergic and dopaminergic pathway gene variants do not influence Huntington’s disease motor onset

Eliana Marisa Ramos · Jeanne C. Latourelle · Tammy Gillis · Jayalakshmi S. Mysore · Ferdinando Squitieri · Alba Di Pardo · Stefano Di Donato · Cinzia Gellera · Michael R. Hayden · Patrick J. Morrison · Martha Nance · Christopher A. Ross · Russell L. Margolis · Estrella Gomez-Tortosa · Carmen Ayuso · Oksana Suchowersky · Ronald J. Trent · Elizabeth McCusker · Andrea Novelletto · Marina Frontali · Randi Jones · Tetsuo Ashizawa · Samuel Frank · Marie-Helene Saint-Hilaire · Steven M. Hersch · Herminia D. Rosas · Diane Lucente · Madaline B. Harrison · Andrea Zanko · Ruth K. Abramson · Karen Marder · James F. Gusella · Jong-Min Lee · Isabel Alonso · Jorge Sequeiros · Richard H. Myers · Marcy E. MacDonald

Received: 26 March 2013 / Accepted: 18 April 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract Huntington’s disease (HD) is a neurodegenerative disorder characterized by motor, cognitive, and behavioral disturbances. It is caused by the expansion of the HTT CAG repeat, which is the major determinant of age at onset (AO) of motor symptoms. Aberrant function of N-methyl-d-aspartate receptors and/or overexposure to dopamine has been suggested to cause significant neurotoxicity, contributing to HD pathogenesis. We used genetic association analysis in 1,628 HD patients to evaluate candidate polymorphisms in N-methyl-d-aspartate receptor subtype genes (GRIN2A rs4998386 and rs2650427, and GRIN2B rs1806201) and functional polymorphisms in genes in the dopamine

E. M. Ramos · T. Gillis · J. S. Mysore · D. Lucente · J. F. Gusella · J. Lee · M. E. MacDonald (✉)
Center for Human Genetic Research, Massachusetts General Hospital, Simchons Research Building, Room 5414, 185 Cambridge Street, Boston, MA 02114 USA
e-mail: macdonal@helix.mgh.harvard.edu

E. M. Ramos · J. Alonso · J. Sequeiros
UnIGENE, IBMC—Institute for Molecular and Cell Biology, University of Porto, Porto, Portugal

J. C. Latourelle · S. Frank · M.-H. Saint-Hilaire · R. H. Myers
Department of Neurology, Boston University School of Medicine, Boston, MA 02118 USA

F. Squitieri · A. Di Pardo
Centre for Neurogenetics and Rare Diseases, IRCCS Neurone, 86077 Pozzilli, IS, Italy

S. Di Donato · C. Gellera
Fondazione IRCCS Istituto Neurologico Carlo Besta, Via Celoria 11, 20133 Milan, Italy

M. R. Hayden
Center for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC V5Z 4H4 Canada

P. J. Morrison
Regional Medical Genetics Centre, Belfast HSC Trust, Belfast BT9 7AB UK

M. Nance
Hennepin County Medical Center, 701 Park Avenue, Minneapolis, MN 55415 USA

C. A. Ross · R. L. Margolis
Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, MD 21287 USA

E. Gomez-Tortosa
Department of Neurology, Fundación Jiménez Díaz, Madrid 28040 Spain

C. Ayuso
Department of Genetics, IIS-Fundación Jiménez Díaz, CIBERER, Madrid 28040 Spain

O. Suchowersky
Department of Medicine, University of Alberta, Edmonton, Alberta T6G 2B7 Canada

R. J. Trent · E. McCusker
Sydney Medical School, University of Sydney, Sydney, NSW 2006 Australia

Published online: 04 May 2013
pathway (DAT1 3′ UTR 40-bp variable number tandem repeat (VNTR), DRD4 exon 3 48-bp VNTR, DRD2 rs1800497, and COMT rs4608) as potential modifiers of the disease process. None of the seven polymorphisms tested was found to be associated with significant modification of motor AO, either in a dominant or additive model, after adjusting for ancestry. The results of this candidate-genetic study therefore do not provide strong evidence to support a modulatory role for these variations within glutamatergic and dopaminergic genes in the AO of HD motor manifestations.

Keywords Huntington’s disease · Glutamate receptors · Dopamine pathway · Genetic modifiers

Introduction

Huntington’s disease (HD) is a dominantly inherited neurodegenerative disorder, usually of adult onset, characterized by involuntary choreiform movements, cognitive impairment, and behavioral changes. HD is caused by the expansion of an unstable polymorphic CAG repeat in HTT [1]. Age at onset (AO) of diagnostic clinical symptoms is inversely correlated with the size of the expanded CAG repeat. It explains about 50–70% of the variance in motor AO [2–4], while the remainder is highly heritable, strongly implying the existence of genetic factors that modulate the rate of the pathogenic process that leads to onset of symptoms [5–7].

The neuropathological changes that comprise the pathological grading system for HD are found in the striatum, where there is a selective and progressive neuronal loss of medium spiny neurons (MSNs) [8, 9]. Glutamatergic and dopaminergic pathways are well known to regulate striatal neuronal function by interacting and modulating each other, suggesting that both glutamate and dopamine receptors may act coordinately in causing deregulation of calcium homeostasis [10–12] with consequent mitochondrial depolarization and caspase activation [13, 14]. Both pathways have been implicated in HD pathogenesis, suggesting that variation in function or expression of glutamate receptor subunits and/or dopamine pathway genes might modulate excitotoxic cell death, thereby modulating AO of symptoms. Indeed, polymorphisms within the genes that encode the NR2A and NR2B glutamate receptors (GRIN2A and GRIN2B) have been implicated in genetic studies with HD patients as potential modifiers of clinical AO [15–18].

Based upon these observations, the aim of the present study was to utilize common and multi-allele functional polymorphisms to test the possibility that genetic variation in genes of the glutamatergic (GRIN2A and GRIN2B) and dopaminergic (COMT, DRD2, DRD4, and DAT1) pathways may explain some of the variation in AO of HD motor manifestations, in a large and well-described cohort of 1,628 HD patients of European ancestry.
Results – Article 6

Material and methods

Subjects We analyzed 1,628 DNA samples from HD patients participating in research from collaborating investigators (HD-MAPS), the HD observational study COHORT, the Harvard Tissue Resource Center Bank (McLean’s Hospital, Belmont, MA) and the National Neurological Research Bank (VAMC Wadsworth Division, Los Angeles, CA). Our cohort comprises a well-described set of HD samples [19] with CAG repeat sizes ranging from 40 to 53 repeats, known motor AO, ancestry, and familial relationship.

Genotyping Repeat sizes of the HTT CAG alleles and DAT1 and DRD4 variable number tandem repeats (VNTRs) were determined using previously established polymerase chain reaction (PCR) amplification assays [20–22]. The size of the products was determined using the ABI PRISM 3730xl automated DNA Sequencer (Applied Biosystems, Foster City, CA) and GeneMapper version 3.7 software. Genotyping of the polymorphisms in GRIN2A (rs4998386 and rs2650427), GRIN2B (rs1806201), COMT (rs4680), and DRD2 (rs1800497) was performed by real-time PCR using commercially available TaqMan Genotyping probes (Applied Biosystems, Foster City, CA) and carried out on the LightCycler® 480 (Roche Diagnostics, Mannheim), following the manufacturer’s instructions.

Statistical analysis Multivariate analyses were conducted using generalized estimating equations (GEE) to assess the association of the different polymorphisms with residual HD motor onset, adjusting for familial component and ancestry. The weighted GEE was computed assuming an independent correlation structure and using the robust estimator of the variance to account for familial relationships. All statistical analyses were performed using PASW Statistics (version 18).

Results

Association with GRIN2A and GRIN2B The genetic evidence supporting a role for GRIN2A or GRIN2B in modulating AO of HD symptoms is equivocal. A candidate gene study with 167 German HD patients reported an association between AO and rs1969060 in GRIN2A and two polymorphisms in GRIN2B (rs1806201 and rs890) [17]. However, in a follow-up study, the same authors found that two other SNPs, rs8057394 and rs2650427, in GRIN2A had a stronger association with AO [16]. A subsequent study in 1,211 European HD individuals found an association of GRIN2A rs2650427, and when stratified by AO subtypes, they found a nominally significant association with rs1969060 (GRIN2A) and rs1806201 (GRIN2B) [15]. On the other hand, in a Venezuelan sample, no evidence was found for the GRIN2B polymorphisms, and a weak association was found for GRIN2A rs1969060 [18]. We attempted to replicate the apparent genetic association of the polymorphisms with the greatest evidence of association with HD AO, namely rs2650427 in GRIN2A and rs1806201 in GRIN2B, as well as an interesting GRIN2A polymorphism associated with decreased Parkinson’s disease (PD) risk in conjunction with heavy coffee consumption (rs4998386) [23]. However, the results of association analysis for each polymorphism failed to demonstrate significant association with HD motor AO (Table 1).

Association with dopamine pathway genes Dopamine pathway genes have not previously been assessed as genetic HD AO modifiers. Therefore, we tested functional polymorphisms in DRD2 and DAT1 believed to affect neurotransmission primarily in the striatum, in addition to polymorphisms in COMT and DRD4, known to have an impact on the frontal cortex function. The Val158Met COMT polymorphism has been shown to affect, in a codominant mode, the activity level of the COMT enzyme that metabolizes dopamine [24–26]. The TaqIA polymorphism in the vicinity of DRD2 is reported to be a genetic marker for D2 receptor density in the brain, with the minor allele being associated with a lower density of this receptor especially in the striatum [27–30]. However, the results of our genetic analysis failed to reveal significant evidence of association of the functional polymorphism in DRD2 or in COMT with HD motor AO (Table 1).

We then evaluated the DRD4 gene, as it has been suggested that different repeat sequences of the multi-allele 48-bp VNTR in DRD4 may differentially affect the gene’s expression and consequently alter D4 receptor density in the brain. The seven-repeat allele had a lower expression compared with two- and four-repeat alleles [31]. Given this observation, our analysis specifically tested the potential association of the seven-repeat allele with motor HD AO. The results demonstrated that the presence of this allele did not explain any variance of AO in our cohort of HD patients (Table 1).

We also assessed the DAT1 gene by evaluating the multi-allele 40-bp VNTR polymorphism. This polymorphism was chosen because it has been reported that individuals with 10/10 repeats have lower dopamine transporter density than individuals with at least one copy of the nine-repeat allele who exhibit more effective dopamine removal at the synapse [32, 33]. However, despite evidence for biological effects, the results of our analysis did not reveal a significant association of the ten-repeat allele with HD motor AO (Table 1).

Discussion

The circuitry of the striatum, where MSNs are particularly vulnerable to the effects of the HD mutation [8, 9], has
Table 1  Multivariate correlation of the polymorphisms in the glutamatergic (GRIN2A and GRIN2B) and dopaminergic (COMT, DRD2, DRD4, and DAT1) pathway genes with residual age at motor onset

<table>
<thead>
<tr>
<th>Gene</th>
<th>Polymorphism</th>
<th>Number of samples</th>
<th>Dominant model</th>
<th>Additive model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Standardized coefficient</td>
<td>p value</td>
</tr>
<tr>
<td>Glutamatergic pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRIN2A</td>
<td>rs4998386</td>
<td>1,585</td>
<td>0.087</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>rs2650427</td>
<td>1,619</td>
<td>-0.014</td>
<td>0.739</td>
</tr>
<tr>
<td>GRIN2B</td>
<td>rs1806201</td>
<td>1,602</td>
<td>-0.056</td>
<td>0.164</td>
</tr>
<tr>
<td>Dopaminergic pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMT</td>
<td>rs4680</td>
<td>1,620</td>
<td>-0.047</td>
<td>0.222</td>
</tr>
<tr>
<td>DRD2</td>
<td>rs1800497</td>
<td>1,625</td>
<td>0.051</td>
<td>0.196</td>
</tr>
<tr>
<td>DRD4</td>
<td>Exon 3 48-bp VNTR</td>
<td>1,527</td>
<td>-0.043</td>
<td>0.322</td>
</tr>
<tr>
<td>DAT1</td>
<td>3’ UTR 40-bp VNTR</td>
<td>1,614</td>
<td>0.062</td>
<td>0.363</td>
</tr>
</tbody>
</table>

*p values were derived using GEE to account for familial relationships and ancestry.

provided a rich source of candidate HD genetic modifiers. Aberrant function of N-methyl-d-aspartate receptors (NMDAR) and overexpression of MSNs to dopamine cause neurotoxicity [34–36], suggesting that variation in expression or function of glutamate receptor subunits and/or dopaminergic pathway genes could modulate excitotoxic death and thereby affect HD AO.

Association of AO with specific polymorphisms in NR2A and NR2B, encoding NMDAR subunits, has been previously reported [15–18]. However, in our sample of European ancestry, we found no definitive evidence of association for either of the two GRIN2A or for the GRIN2B SNPs that were tested with the residual AO after accounting for the effect of Htt CAG repeat length. One SNP, rs1806201 in GRIN2B, was close to nominal significance (p=0.053 in the additive model). Though this value would not survive correction for the multiple hypotheses tested in our study, it may be of interest for future modifier studies given the effects previously reported [15–17]. The lack of replication in our sample of the reported associations might be explained by different study designs, including the patient populations and definition of the phenotypic trait. We have previously shown that stringent sample selection and analysis criteria are critical factors in HD association studies. Indeed, genetic background related to ancestry [19] and non-normal distribution of CAG allele size [37] can have a profound confounding effect when testing for the effects of potential genetic modifiers.

Our test of the GRIN2A rs4998386 polymorphism that has been recently associated with decreased risk of developing PD in individuals who are heavy coffee drinkers [23] was an attempt to assess a neurodegenerative disease–associated risk allele that may interact with a common environmental factor. We did not find evidence of association of this particular SNP with AO of HD motor symptoms. Though coffee consumption data are not available on our study subjects, this negative result is consistent with previous genetic findings showing that the Htt CAG repeat polymorphism is not a modifier of PD onset [38], strongly suggesting that the pathogenic process that culminates in HD manifestations may be distinct from the neurodegenerative disease process that leads to PD symptoms.

Genes in the dopaminergic pathway have not previously been evaluated as potential modifiers of the AO of overt motor symptoms in HD. We selected polymorphisms in four dopaminergic pathway genes that have been investigated in other neurological disorders because they are believed to affect neurotransmission by affecting the level of the enzyme that metabolizes dopamine, the density of dopamine receptors, and the activity of dopamine transporter. Our results did not reveal a significant modifying effect of any of the four polymorphisms, in COMT, DRD2, DRD4, or DAT1, on the onset of HD motor symptoms and therefore fail to support a role for these functional variants in the disease process that leads to the onset of neurological symptoms in HD.

In summary, the results of our study did not provide evidence for an association of DNA variants that affect the biology of particular genes in the glutamatergic pathway (GRIN2A, GRIN2B) or the dopaminergic pathway (DRD2, DRD4, DAT1, COMT) with the Htt CAG repeat length–dependent disease process that leads to the onset of clinical motor manifestations of HD. However, our study does not preclude the possibility that other DNA variants in these genes, or in other genes involved in these pathways, may act as genetic AO modifiers. Moreover, it is also possible that genes in the glutamatergic and dopaminergic pathways may serve to modify the rate of progression of the distinct processes that determine the rate of decline in the ∼18-year period between the age at clinical diagnosis of HD and death, which is independent of the size of the Htt CAG repeat [39].
Acknowledgments The authors thank Drs. Ignacio Muñoz-Sanjuan, Seung Kwak, and Jamshed Arjomand from CHDI Foundation, Inc., as well as the HD families, whose participation in genetic studies made this work possible; the COHORT co-investigators and contributors (see "Appendix"); contributors to the HD-MAPS study, Drs. Alexandra Durr, Adam Rosenblatt, Luigi Frati, Susan Perlman, P. Michael Connelly, Mary Lou Klimek, Melissa Diggin, Tiffany Hadzi, and Ayana Duckett; the Harvard Brain Tissue Resource Center at McLean Hospital; and the National Neurological Research Specimen Bank at the VA West Los Angeles Healthcare Center. EM8 is the recipient of a scholarship from the Fundação para a Ciência e a Tecnologia of Portugal (SFRH/BD/44335/2008). Study funding was supported by NIH grants from the NINDS NS16367 (The Massachusetts HD Center Without Walls) and NS32765, and the CHDI Foundation, Inc.

Ethical standards This study used only deidentified, previously collected DNA samples and phenotypic data in a manner approved by the Institutional Review Board of Partners HealthCare, Inc.

Conflict of Interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Appendix

Huntington Study Group COHORT investigators

Steering committee

University of Rochester, Rochester, NY: Ira Shoulson, MD (principal investigator), James Gusella, PhD (co-principal investigator), Tatiana Foroud (co-principal investigator), Irina Antonijevic, Dan van Kammen (CHDI Foundation Inc.)

Publications and data use committee

Tatiana Foroud (Chair), Ray Dorsey (Co-Chair), John Warner (CHDI), Joe Giuliano, Louise Vetter, Oksana Suchoversky; Christopher Beck, David Oakes

Participating investigators and coordinators

F. Marshall, University of Rochester; K. Marder, S. Frucht, C. Moskowitz, R. Clouse, P. Wasserman, Columbia University Medical Center; K. Shannon, J. Jaglin, Rush University Medical Center; J. Jankovic, A. Palao, Baylor College of Medicine; M. Harrison University of Virginia; C. Singer, M. Quesada, University of Miami; S. Hersch, D. Rosas, K. Tanev, K. Malarick, Massachusetts General Hospital;

A. Colehner University of Pennsylvania; J. Sanchez-Ramos, University of South Florida; S. Kostyk, Ohio State University; J. Paulsen, University of Iowa; J. Perlmuter, S. Tabbal, Washington University; C. Ross, F. Nucifora, Johns Hopkins University; R. Dubinsky, H. Dubinsky University of Kansas Medical Center; O. Suchoversky, M.L. Klimek, University of Calgary; R. Jones, C. Testa, S. Factor, Emory University School of Medicine; D. Jennings, Institute for Neurological Disorders; J. Morgan Medical College of Georgia; D. Higgins, E. Mohlo, Albany Medical College; J. Adams, The Centre for Addiction and Mental Health, Toronto, Canada; S. Frank, M. Saint-Hilaire, M. Diggin, Boston University; S. Furtado, University of Alberta; F. Walker, C. O’Neill, V. Hunt, Wake Forest University School of Medicine; K. Quaïd, Indiana University School of Medicine; M. LeDoux, University of Tennessee Health Science Center; L. Raymond, B. Leavitt, J. Decolongon, University of British Columbia, Canada; S. Perlman, University of California, Los Angeles; J. Corey-Bloom, G. Peavy, J. Goldstein, University of California San Diego; R. Kumar, Colorado Neurological Institute; E. McCusker, J. Griffith, C. Loy, Westmead Hospital, NSW, Australia; V. Wheelock, T. Tempkin, A. Martin, University of California Davis; M. Nance, Hennepin County Medical Center; U. Kang, University of Chicago; W. Mallonee, G. Suter, Hereditary Neurological Disease Center, Kansas; F. Revilla, M. Gartner University of Cincinnati/Cincinnati Children’s Hospital; C. Drazin, M.J. Fitzpatrick, University of Connecticut; M. Pamisset, Hôtel-Dieu Hôpital-CHUM, Canada; K. Duff, University of Utah; B. Scott, Duke University Medical Center; W. Weiner, B. Robotson, University of Maryland School of Medicine; E. Chia, O. Yastrubetskaya, A. Churchyard, St Vincent’s Aged Mental Health Service (SVAMHS) Melbourne, Australia; T. J. Greenamyre; University of Pittsburgh; P. Agarwal, Booth Gardner Parkinson’s Care Center, WA; Biostatistics/Coordination Center: University of Rochester—D. Oakes, C. Beck, S. Robertson, K. Eaton, P. Lindsay, L. Deuel;

© Springer
DNA Genotyping Center: M. MacDonald, Center for Human Genetic Research, Massachusetts General Hospital

Contributors
C. Hickey, University of Rochester; L. Muratori, E. Louis, Columbia University Medical Center; A. Leserman, N. Doucette, Ohio State University; E. Ulc, R. Rodnitzky, S. Vik, University of Iowa, R. Davis, S. Dietrich, University of Virginia; V. Segro, D. Erickson, Colorado Neurological Institute; V. Hunt, Wake Forest University School of Medicine; N. Lucarelli, University of Pittsburgh; J. Broyles, University of Kansas Medical Center; J. DeLaRosa, University of Texas (Galveston); E. Louis, Columbia University Medical Center; P. Panegyres, Neurodegenerative Disorders Research, WA, AUS; A. Schmidt, S. Barton, Washington University; E. Sperin, C. Testa, Emory University; F. Thiede, University of California, Los Angeles; S. E. Zauber, M. Wesson, Indiana University; R. McInnis, Massachusetts General Hospital; C. Welsh, Johns Hopkins University; M. Wesson, Indiana University School of Medicine; A. Coleman, University of British Columbia, Canada

References

Genetic modifiers of Huntington’s disease
Article 7: Prevalence of other polyglutamine repeat expansions in Huntington’s disease patients.

Ramos EM, Gillis T, Mysore JS, Lee JM, Sequeiros J, Alonso I, Gusella JF on behalf of the Huntington Study Group COHORT Investigators, MacDonald ME.

[in preparation]
Prevalence of other polyglutamine repeat expansions in Huntington’s disease patients

Eliana Marisa Ramos¹,², Tammy Gillis¹, Jayalakshmi S. Mysore¹, Jong-Min Lee¹, Jorge Sequeiros²,³,⁴, Isabel Alonso²,³,⁴, James F. Gusella¹ on behalf of the Huntington Study Group COHORT Investigators, Marcy E. MacDonald¹

¹ Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; ² UnIGENe, IBMC – Institute for Molecular and Cell Biology, University of Porto, Porto, Portugal; ³ CGPP, IBMC – Institute for Molecular and Cell Biology, University of Porto, Porto, Portugal; ⁴ ICBAS – Instituto de Ciências Biomédicas Abel Salazar, University of Porto, Porto, Portugal

Corresponding Author M.E. MacDonald, Center for Human Genetic Research, Massachusetts General Hospital, Simches Research Building, Room 5414, 185 Cambridge Street, Boston, MA 02114, USA
Telephone: +1(617)-726-5089
Fax: +1(617)-726-5735
Email: macdonam@helix.mgh.harvard.edu

Running title PolyQ disease genes in HD patients

Abstract Huntington’s disease (HD) is a dominantly inherited neurodegenerative disorder caused by an abnormal extended polyglutamine stretch in the huntingtin protein. At least eight additional neurodegenerative disorders have been associated with expanded polyglutamine domains in different proteins. The allele size responsible for these so called polyglutamine disorders have been pointed out as possible pathogenesis contributors in diverse neurodegenerative disorders. To evaluate the contribution of other expanded polyglutamines to the etiology of HD, we determined the CAG length of six other polyglutamine disease genes known to cause different forms of spinocerebellar ataxia (SCA) - ATXN1, ATXN2, ATXN3, CACNA1A, TBP and ATN1 - in a large and well-characterized cohort of 1471 Caucasian individuals with HD. We found seven HD cases
that carried intermediate SCA2 alleles, known to be associated with parkinsonism and amyotrophic lateral sclerosis in addition to cerebellar ataxia, and extremely rare in the general population. Three cases carried reduced to fully penetrant SCA17 alleles. Otherwise, CAG allelic distribution did not differ between HD cases and controls. Though our results do not support an obvious genetic association between HD and CAG repeat length in the investigated genes, it cannot be ruled out that, in some individuals with HD, SCA2 and SCA17 disease-associated alleles may contribute to disease phenotype.

**Keywords** Huntington’s disease; polyglutamine disorders; Neurodegenerative disorders; Triplet repeats

**Introduction** Huntington’s disease (HD) is one of at least nine dominantly inherited neurodegenerative disorders caused by the abnormal expansion of an unstable polymorphic CAG repeat in the translated region of disease-specific genes, resulting in the production of proteins with expanded polyglutamine (polyQ) domains. This family of disorders, commonly known as polyQ diseases, is characterized by a constant progression of symptoms that usually appear later in life, with age-at-onset and severity being inversely correlated with the size of the CAG repeat\(^1\). Even though both normal and mutant proteins are ubiquitously expressed, neurons are extremely sensitive to the progressive accumulation of mutant proteins causing specific dysfunction and eventually death of specific types of neurons in each polyQ disease\(^2\). These shared features suggest a common toxic effect related to the expanded polyQ sequence that has an important role on the mechanisms leading to the disease. The exact mechanism of pathogenesis remains uncertain, but several mechanisms have been proposed for the pathogenesis of polyQ diseases, including cleavage into toxic fragments, protein misfolding and aggregation and dysregulation of transcription\(^2\). Another possible mechanism involves the alteration of the protein conformation due to the expanded polyQ, resulting in aberrant protein-protein interactions. These aberrant interactions may include interactions of the expanded polyQ with other polyQ proteins, as it has been shown that polyQ proteins have higher tendency to interact with other polyQ proteins\(^3\). Despite all the common aspects, each polyQ expansion causes different neurodegeneration and phenotypic profiles as the disease-causing proteins share no homology outside the polyQ domain and have different subcellular localization, abundance, structure and function\(^4\).

In the present study, we sought to determine whether expansions in other polyQ disease genes were associated with HD and could contribute to the disease. We tested the possible effect of some analogous polyQ expansions in HD by determining the CAG
repeat length of six additional polyQ disease genes: ATXN1, ATXN2, ATXN3, CACNA1A, TBP and ATN1 in a large and well-characterized cohort of HD patients.

Material and Methods

Subjects. This study comprised a total of 1471 HD cases and 609 “non-HD” control subjects from the observational study The Cooperative Huntington Observational Research Trial (COHORT). This is a cohort study conducted on members of families affected by HD in the United States, Canada and Australia, which collected demographic and clinical data on all participants. All subjects included in this study were unrelated and self-reported as Caucasian.

Genotyping. CAG repeat lengths for each polyQ disease gene were determined by polymerase chain reaction (PCR) assays, using fluorescently labelled primers, with minor modifications. PCR products were run on an ABI PRISM 3730XL automated DNA Sequencer (Applied Biosystems, Foster City, CA) and analysed using GeneMapper version 3.7 software. A set of genomic DNA standards was sequenced for each polyQ repeat and used to provide CAG allele size standards. Expansions above the normal repeat range were sequenced after gel separation to further confirm the number of CAGs and the presence of CAA/CAT interruptions.

Statistical analysis. Two-tailed Fisher’s exact test was used to compare each polyQ disease gene CAG repeat frequency between HD and control chromosomes. The statistical analyses were performed using PASW Statistics 18 (SPSS Inc., Chicago, IL).

Results. To evaluate the potential contribution of other polyQ repeats in HD, we determined the CAG repeat length in six polyQ genes associated with spinocerebellar ataxias (SCA): ATXN1, ATXN2, ATXN3, CACNA1A, TBP and ATN1. Both HD and control chromosomes carried CAG repeats within the non-pathological range for mostly all the polyQ disease genes, with the exception of ATXN2 and TBP (Table 1).

The ATXN2 gene contains a variable CAG repeat with non-pathological alleles ranging from 14 to 31 triplets (22-23 alleles account for about 98% of normal alleles) and expansions of more than 34 triplets causing SCA2. ATXN2 alleles with 32 to 34 CAGs are considered intermediate, with reduced penetrance and a variable expressivity. Seven out of 1463 HD patients had an ATXN2 allele in the intermediate range: one with 32 repeats and six with 33 repeats, accounting for 0.55% of the HD cohort. Only one out of 608 (0.16%) control individuals carried an intermediate ATXN2 allele with 33 CAGs. Sequencing of the eight alleles with >32 CAGs showed that they were all interrupted with
one CAA triplet. Despite the fact that the prevalence of intermediate \textit{ATXN2} alleles is numerically higher in HD patients, we did not observed significant differences in the \textit{ATXN2} polyQ lengths between HD cases and controls (Table 1).

In the \textit{TBP} gene, the normal CAG repeat range reported is from 25 up to 42 triplets, while alleles with 49 or more repeats are fully penetrant and are known to cause SCA17 and expansions between 43 and 48 repeats show incomplete penetrance. Among the 1126 cases with HD genotyped for the \textit{TBP} CAG repeat, three individuals (0.27%) carried pathological SCA17 alleles: two within the reduced penetrance range (43 and 45 triplets) and one fully penetrant allele (49 triplets). On the other hand, we only observed one reduced penetrance allele (44 triplets) among 420 control individuals. We did not observed significant differences in the \textit{TBP} polyQ lengths between HD cases and controls (Table 1).

\textbf{Discussion} The phenotypic and molecular overlap that exists between HD and other polyQ disorders has implied that these family of disorders may share common pathogenic processes initiated by DNA repeat mutations at distinct genetic loci. Our current study aimed at determine the prevalence of expanded CAG repeats in polyQ disease genes in a well-characterized cohort of individuals diagnosed with HD, to assess whether these alleles may, or may not, contribute to the variable phenotypes (expressivity) and/or age-at-onset of HD. Though there wasn’t a significant difference in the allele distribution for the loci assessed, we found ten HD cases in our cohort that carried expanded CAG alleles in two other polyQ disease genes, SCA2 and SCA17.

SCA2 is mainly characterized by gait and limb ataxia and a marked slowing of saccadic eye movement, that are caused by a pure CAG repeat of more than 34 triplets. Borderline \textit{ATXN2} expansions interrupted with CAA triplets were observed as the cause of Levodopa-responsive Parkinson’s disease\textsuperscript{7} and have more recently been associated with increased risk for amyotrophic lateral sclerosis (ALS)\textsuperscript{8-10} and tauopathy progressive supranuclear palsy\textsuperscript{11}. We have found seven HD cases that carried \textit{ATXN2} expansions of 32-33 repeats and even though there wasn’t a significant difference in the frequency of this range of alleles between our HD cases and controls, it has been previously shown that these intermediate repeats are extremely rare in control subjects (3 out of 4877 healthy Caucasian control subjects\textsuperscript{11}). All the intermediate SCA2 alleles observed in our HD cases had one CAA interruption, as it was described for intermediate alleles associated with parkinsonism and ALS. Most of these HD cases presented, at baseline, mild to moderate slowing of the saccade velocity, mild dysarthria and moderate to severe
gait impairment, clinical features that overlap with SCA2 phenotype. Furthermore, one of these cases also presented bilateral rest tremor and drug induced parkinsonism.

We have also found, in our cohort, three HD cases that carried expanded SCA17 alleles. This SCA is typically characterized by ataxic gait and dysarthria, with personality changes, depression, cognitive impairment, epilepsy and chorea\(^{12}\). Hence, SCA17 is also classified as a HD-like syndrome. It is caused by CAG/CAA repeat expansions in the gene encoding the TATA-box binding protein (TBP), which is an important transcription initiation factor. Previous studies, have shown that both TBP and huntingtin proteins are part of the insoluble protein fraction from postmortem HD brains, implying a potential functional overlap of these proteins\(^{13,14}\). It is possible, therefore, that these individuals, with both the HTT CAG repeat expansion and the SCA17 CAG expansion, may have a distinct clinical presentation, compared to typical HD individuals, without SCA17 expanded alleles. Unfortunately, it was not possible to obtain detailed clinical information to evaluate this possibility. But these results encourage further assessment of SCA17 in HD patient cohorts.

Our results do not support an evident genetic association between HD and expanded CAG repeats in the polyQ disease genes we analyzed. We did, however, observe that a small proportion of HD patients carried SCA2 intermediate alleles, reinforcing common pathways resulting in polyQ-mediated neurodegeneration. Therefore, further studies examining the prevalence of intermediate SCA2 alleles in a larger cohort of HD and healthy controls and its association with different HD features, such as age-at-onset, may help uncover a possible effect of the CAG repeat at ATXN2 on the etiology of HD.

**Acknowledgements** The authors thank Drs. Ignacio Muñoz-Sanjuan, Seung Kwak and Jamshid Arjomand from CHDI Foundation, Inc. and the COHORT co-investigators and contributors (see “Appendix”), as well as the HD families, whose participation in genetic studies made this work possible. EMR is the recipient of a scholarship from the Fundação para a Ciência e a Tecnologia of Portugal (SFRH/BD/44335/2008). Study funding supported by NIH grants from the NINDS NS16367 (The Massachusetts HD Center Without Walls) and NS32765, and the CHDI Foundation, Inc.

**Funding Sources** EMR is the recipient of a scholarship from Fundação para a Ciência e a Tecnologia (SFRH/BD/44335/2008). Grants from the National Institutes of Health NINDS NS16367 Huntington’s Disease Center Without Walls (JFG, MEM), NS32765 (MEM) and the CHDI Foundation, Inc. (JFG, MEM).
Conflict of Interests The authors declare no conflict of interest.

References

### Table

#### Table 1. Prevalence of other polyQ disease repeats in Huntington’s disease and control subjects.

<table>
<thead>
<tr>
<th>PolyQ gene</th>
<th>Huntington’s Disease</th>
<th>Controls</th>
<th>Differences in polyQ length in HD vs. control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PolyQ length observed</td>
<td>Number of chromosomes</td>
<td>PolyQ length observed</td>
</tr>
<tr>
<td>ATXN1 (SCA1)</td>
<td>21-38 CAGs&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2328</td>
<td>25-38 CAGs&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>ATXN2 (SCA2)</td>
<td>13-31 CAGs&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2919</td>
<td>13-31 CAGs&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>ATXN3 (SCA3)</td>
<td>14-41 CAGs&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2350</td>
<td>14-44 CAGs&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>CACNA1A (SCA6)</td>
<td>4-15 CAGs&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2302</td>
<td>4-15 CAGs&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>TBP (SCA17)</td>
<td>25-42 CAGs&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2252</td>
<td>27-41 CAGs&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>ATN1 (DRPLA)</td>
<td>4-25 CAGs&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2298</td>
<td>6-25 CAGs&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Abbreviations: PolyQ, polyglutamine; SCA, spinocerebellar ataxia; DRPLA, dentatorubral-pallidoluysian atrophy. <sup>a</sup>Normal, <sup>b</sup>Intermediate, <sup>c</sup>Reduced penetrance and <sup>d</sup>Full penetrance alleles
Appendix

Huntington Study Group COHORT Investigators

Steering Committee:

*University of Rochester, Rochester, NY:* Ira Shoulson, MD (principal investigator), James Gusella, PhD (co–principal investigator), Tatiana Foroud (co-principal investigator), Irina Antonijevic, Dan van Kammen (CHDI Foundation Inc.)

Publications & Data Use Committee:

Tatiana Foroud (Chair), Ray Dorsey (Co-Chair), John Warner (CHDI), Joe Giuliano, Louise Vetter, Oksana Suchowersky, Christopher Beck, David Oakes

Participating Investigators and Coordinators:

F. Marshall, University of Rochester;

K. Marder, S. Frucht, C. Moskowitz, R. Clouse, P. Wasserman, Columbia University Medical Center;

K. Shannon, J. Jaglin, Rush University Medical Center;

J. Jankovic, A. Palao, Baylor College of Medicine;

M. Harrison University of Virginia;

C. Singer, M. Quesada, University of Miami;

S. Hersch, D. Rosas, K. Tanev, K. Malarick, Massachusetts General Hospital;

A. Colcher University of Pennsylvania;

J. Sanchez-Ramos, University of South Florida;

S. Kostyk, Ohio State University;

J. Paulsen, University of Iowa;

J. Perlmutter, S. Tabbal, Washington University;

C. Ross, F. Nucifora, Johns Hopkins University;

R. Dubinsky, H. Dubinsky University of Kansas Medical Center;

O. Suchowersky, M.L. Klimek, University of Calgary
R. Jones, C. Testa, S. Factor, Emory University School of Medicine;  
D. Jennings, Institute for Neurological Disorders  
J. Morgan Medical College of Georgia;  
D. Higgins, E. Mohlo, Albany Medical College;  
J. Adams, The Centre for Addiction and Mental Health, Toronto, Canada;  
S. Frank, M. Saint-Hilaire, M. Diggin, Boston University;  
S. Furtado, University of Alberta  
F. Walker, C. O'Neill, V. Hunt, Wake Forest University School of Medicine;  
K. Quaid, Indiana University School of Medicine;  
M. LeDoux, University of Tennessee Health Science Center;  
L. Raymond, B. Leavitt, J. Decolongon, University of British Columbia, Canada;  
S. Perlman, University of California, Los Angeles  
J. Corey-Bloom, G. Peavy, J. Goldstein, University of California San Diego;  
R. Kumar, Colorado Neurological Institute;  
E. McCusker, J. Griffith, C. Loy, Westmead Hospital, NSW, Australia;  
V. Wheelock, T. Tempkin, A. Martin, University of California Davis;  
M. Nance, Hennepin County Medical Center;  
U. Kang, University of Chicago;  
W. Mallonee, G. Suter, Hereditary Neurological Disease Center, Kansas;  
F. Revilla, M. Gartner University of Cincinnati/Cincinnati Children's Hospital;  
C. Drazinic, M.J. Fitzpatrick, University of Connecticut;  
M. Panisset, Hôtel-Dieu Hôpital-CHUM, Canada;  
K. Duff, University of Utah  
B. Scott, Duke University Medical Center;  
W. Weiner, B. Robottom, University of Maryland School of Medicine;  
E. Chiu, O. Yastrubetskaya, A. Churchyard, St Vincent's Aged Mental Health Service (SVAMHS) Melbourne, Australia;  
T. J. Greenamyre; University of Pittsburgh;  
P. Agarwal, Booth Gardner Parkinson's Care Center, WA

Biostatistics/Coordination Center:
University of Rochester – D. Oakes, C. Beck, S. Robertson, K. Eaton, P. Lindsay, L. Deuel

DNA Genotyping Center:

M. MacDonald, Center for Human Genetic Research, Massachusetts General Hospital

Contributors:

C. Hickey, University of Rochester; L. Muratori, E. Louis, Columbia University Medical Center; A. Leserman, N. Doucette, Ohio State University; E. Uc, R. Rodnitzky, S. Vik, University of Iowa, R. Davis, S. Dietrich, University of Virginia; V. Segro, D. Erickson, Colorado Neurological Institute; V. Hunt, Wake Forest University School of Medicine; N. Lucarelli, University of Pittsburgh; J. Broyles, University of Kansas Medical Center; J. DeLaRosa, University of Texas (Galveston); E. Louis, Columbia University Medical Center; P. Panegyres, Neurodegenerative Disorders Research, WA, AUS; A. Schmidt, S. Barton, Washington University; E. Sperin, C. Testa, Emory University; F. Thiede, University of California, Los Angeles; S. E. Zauber, M. Wesson, Indiana University; R. McInnis, Massachusetts General Hospital; C. Welsh, Johns Hopkins University; M. Wesson, Indiana University School of Medicine; A. Coleman, University of British Columbia, Canada
**Huntington's Disease genetic interactions**

HD can cause markedly different patterns of symptoms, including a wide range of prominent cognitive and psychiatric disturbances, such that some patients may have initially been diagnosed with depressive disorder, BP or schizophrenia. Factors modulating the disease mechanism that leads to the onset of these cognitive and psychiatric symptoms are probably different from those that are involved in the occurrence of chorea and other motor manifestations.

Recent findings by our group demonstrated an exciting genetic overlap of major depressive disorder (MDD) with the HTT CAG triplet repeat, strongly supporting the existence of shared biological underpinnings of HD with major psychiatric and neurodevelopmental disorders. We have now tested the potential overlap of the biology that is modulated by HTT CAG repeat with the complex biology that determines ALS and BP, by determining the prevalence of high-end normal/reduced penetrance HTT CAG repeats in large cohorts of patients.

Our results showed that the allele frequencies in sporadic ALS subjects, speculated to share with HD similar neurodegenerative pathways, though not clinical symptoms, were not significantly different from the non-HD chromosomes and that the HTT CAG repeat has no effect on susceptibility to ALS or to its phenotypic manifestations [Article 8]. Likewise, the distribution of HTT CAG repeat alleles in DSM-IV diagnosed cases of BP patients was similar to that of control individuals screened for psychiatric disease, with an allele distribution at the high-end normal/reduced penetrance allele range that can be explained by the frequency of major depression in the general population [Article 9]. Moreover, our results provided important estimates of the frequency of alleles in the general population that may give rise to new HD families.
Article 8: Prevalence of Huntington’s disease gene CAG repeat alleles in sporadic amyotrophic lateral sclerosis patients.


Reprinted with the permission of Informa Healthcare
License number 3194381148325
Prevalence of Huntington’s disease gene CAG repeat alleles in sporadic amyotrophic lateral sclerosis patients

ELIANA MARISA RAMOS1,2, PAMELA KEAGLE3, TAMMY GILLIS1, PATRICK LOWE3, JAYALAKSHMI S. MYSORE1, ASHLEY LYN LECLERC3, ANTONIA RATTI3, NICOLA TICOZZI3, CINZIA GELLERA3, JAMES F. GUSSELLA1, VINCENZO SILANI1, ISABEL ALONSO2, ROBERT H. BROWN JR3, MARCY E. MACDONALD1 & JOHN E. LANDERS3

1Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA, 2Unigen, IBMC – Institute for Molecular and Cell Biology, University of Porto, Porto, Portugal, 3Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts, USA, 4Department of Neurology and Laboratory of Neurosciences, ‘Dino Ferrari’ Center, Universita degli Studi di Milano – IRCCS Istituto Neurologico, Milan, Italy and 5Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, Milan, Italy

Abstract
A higher prevalence of intermediate ataxin-2 CAG repeats in amyotrophic lateral sclerosis (ALS) patients has raised the possibility that CAG expansions in other polyglutamine disease genes could contribute to ALS neurodegeneration. We sought to determine whether expansions of the CAG repeat of the HTT gene that causes Huntington's disease, are associated with ALS. We compared the HTT CAG repeat length on a total of 3144 chromosomes from 1572 sporadic ALS patients and 4007 control chromosomes, and also tested its possible effects on ALS-specific parameters, such as age and site of onset and survival rate. Our results show that the CAG repeat in the HTT gene is not a risk factor for ALS nor modifies its clinical presentation. These findings suggest that distinct neuronal degeneration processes are involved in these two different neurodegenerative disorders.

Key words: Huntington’s disease, amyotrophic lateral sclerosis, trinucleotide repeat, neurodegenerative disease, polyglutamine expansion

Introduction
Huntington’s disease (HD) is a progressive neurodegenerative disorder characterized by motor, cognitive and behavioral manifestations. It is caused by the expansion of an unstable polymorphic CAG repeat, in the first exon of the HTT gene, on chromosome 4p16.3 (1), resulting in an expanded polyglutamine (polyQ) tract in the huntingtin protein. Alleles with fewer than 35 CAGs are considered to produce no symptoms of HD and can be stable (alleles smaller than 27 CAGs) or prone to expansion (27–35 CAGs). A case report raises the question of whether, in rare instances, these high-end ‘normal’ alleles may lead to HD symptoms (2). A third class of alleles between 35 and 39 repeats shows incomplete HD penetrance while alleles with 40 repeats or more are fully penetrant.

Amyotrophic lateral sclerosis (ALS) is an adult onset disorder caused by the loss of motor neurons in the motor cortex, brainstem and spinal cord. The motor neuron loss results in muscle weakness, wasting, fasciculations, spasticity and hyperreflexia, leading to generalized paralysis and death (two to five years after disease onset). The disease is mostly sporadic (SALS) but ~10% of patients have a family history (FALS). The most frequent mutations are found in the superoxide dismutase-1 gene (SOD1), which accounts for 12–23% of FALS (3), in the TAR DNA binding protein 43 (TDP-43) gene (TARDBP) (4,5) and the fused-in sarcoma/translocated in liposarcoma gene (FUS) (6,7), which equally account for ~5% of FALS. Both FUS and TDP-43 proteins have been found in HD inclusions (8–10). Mutations in the OPTN (optineurin) gene have recently been identified.
in a small number of ALS cases (11,12). Neurons from HD patients occasionally display granular cytoplasmic OPTN immunopositivity (13). Huntingtin protein also interacts with the optineurin/Rab8 complex and can influence post-Golgi trafficking to lysosomes (13,14).

Very recently, an expanded hexanucleotide repeat (GGGGCC) within Osym72 has been identified in approximately 23.5%–46% of FALS cases (15,16). Similarly, a genetic overlap has been reported between ALS and the neurodegenerative disorder spinocerebellar ataxia-2 (SCA2). Expansions of the CAG repeat in the ATXN2 gene, which cause SCA2, have been associated with increased risk of SALS (17–19). In this study, we sought to test whether an association exists between the CAG repeat in the HTT gene and SALS phenotypes, focusing on disease susceptibility, age at onset, bulbar versus spinal onset and survival.

Materials and methods

ALS patients and control samples

The patient cohort comprised, in total, 1572 ALS patients. The largest set of samples included 1227 Caucasian SALS cases with known age at onset, which ranged from 19 to 93 years. The mean age at onset was 55.42 years. From these samples, 1200 subjects had available information regarding bulbar or spinal onset, while survival information was known for the 601 deceased SALS individuals. A second set of 345 Italian SALS samples was added to this study. Age at onset, which showed a mean of 57.12 years, was known for 308 of these subjects. Bulbar or spinal onset information was available for 291 individuals while survival information was known for 258.

A large cohort of patients with HD (4007) was used for comparison. The normal chromosomes of these individuals (the one with the shorter non-expanded repeat, i.e. with less than 35 HD CAG repeats) were used for comparison in the CAG repeat frequency analysis, as they should reflect the population distribution of the normally polymorphic HTT CAG allele (20).

Genotyping

The HTT CAG repeat length was determined by a polymerase chain reaction amplification assay, using fluorescently labelled primers, as described (21). The size of the fragments was determined using the ABI PRISM 3730XL automated DNA Sequencer (Applied Biosystems, Foster City, California) and GeneMapper version 3.7 software. A set of HTT CAG alleles, whose lengths were confirmed by DNA sequencing, was used to provide size standards.

Statistical analysis

Fisher’s exact test was used to compare HTT CAG repeat frequency between the SALS cases and comparison non-HD chromosomes. The HTT CAG distribution among these two groups was also analyzed by use of the Mann-Whitney U test, since the samples did not display a normal distribution. To assess the possible effect on bulbar and spinal onset subgroups, Fisher’s exact test was used to compare frequency of HTT alleles between patients in these subgroups. The analysis of correlation of CAG repeat size and
Genetic modifiers of Huntington’s disease

Results

Table I. HD CAG repeat alleles in sporadic ALS.

<table>
<thead>
<tr>
<th>Allele Description</th>
<th>Normal HD allele (&lt; 27CAGs)</th>
<th>High normal HD allele (27-35CAGs)</th>
<th>At risk HD allele (&gt; 35CAGs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control chromosomes</td>
<td>3869 (96.56%)</td>
<td>138 (3.44%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>SALS chromosomes</td>
<td>3044 (96.82%)</td>
<td>99 (3.15%)</td>
<td>1 (0.03%)</td>
</tr>
<tr>
<td>Bulbar onset</td>
<td>729 (97.28%)</td>
<td>21 (2.80%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Spinal onset</td>
<td>2156 (96.59%)</td>
<td>75 (3.30%)</td>
<td>1 (0.04%)</td>
</tr>
</tbody>
</table>

Survival and age at onset was performed using linear regression analysis. A second regression model was performed using logarithmically transformed age at onset and survival, similar to Andresen et al. (22). The statistical analyses were performed using PASW Statistics 18 and GraphPad Prism 5.

Results

We compared the HTT CAG repeat length in genomic DNA from a total of 3144 chromosomes from 1572 sporadic ALS patients and 4007 control non-HD chromosomes (Figure 1). The HTT CAG alleles were divided into three groups: normal (<27 repeats), high normal (27–35 repeats) and low and high HD penetrance or ‘at risk’ (>35 repeats).

In controls, the repeat length varied between nine and 34 HTT CAGs, while the alleles in the SALS sample ranged from eight to 38 repeats, with three alleles having 35 CAGs, the top of the high-end normal range, and a single 38 CAG allele, in the reduced penetrance HD range. The allele frequencies in the SALS subjects were not significantly different from the control chromosomes in either the comparison of normal versus high normal alleles (Fishers exact test, $p = 0.506$) or normal versus >26 CAGs (Fishers exact test, $p = 0.551$) (Table I). The HTT CAG distribution was also analyzed using additional statistical methods to assess skewness (1.115 ± 0.039 for control and 1.173 ± 0.044 for SALS chromosomes). Since the samples were not normally distributed, the Mann-Whitney U test was used to compare the two groups of chromosomes. It showed no significant difference between the control and SALS chromosomes with respect to HTT CAG distribution (Mann-Whitney U test, $p = 0.233$).

Also, we found no association between HTT CAG repeat length and site of ALS onset, as the SALS patients with bulbar onset did not differ significantly from those with spinal onset either in the frequency of normal length versus high normal alleles (Fishers exact test, $p = 0.350$) or in the frequency of normal length alleles (<27CAGs) versus all alleles >26 CAGs (Fishers exact test, $p = 0.476$) (Table I).

The age at onset for ALS manifestations was known for 1535 samples that were successfully genotyped for HTT CAG repeat length and the potential relationship between these two measures was tested by two different linear regression methods. In the first, the onset age (years) for each subject was plotted versus their larger HTT CAG allele, while in the second the natural log of onset age was used (data not shown). Even though a slight negative slope was observed ($-0.06941 \pm 0.09446$ and $-0.0007564 \pm 0.001868$) with both methods, neither reached statistical significance ($p = 0.4626$ and $p = 0.6856$, respectively).

Survival information was known for 859 deceased SALS samples and was similarly analyzed with respect to HTT CAG length by two methods. In an initial analysis, with 601 SALS cases for which survival information was available, we observed a nominally significant association ($p = 0.0235$ ignoring any multiple testing correction) between the HTT CAG repeat length and SALS survival (years) (Figure 2 (left)). This suggested a possible correlation between survival and HTT CAG allele length, with a decrease of ~0.90 months for each additional CAG repeat. In order to further explore this finding, we tested whether HTT CAG repeat length had an effect on survival in an additional population of 258 Italian SALS patients. However, the Italian SALS patients’ survival data did not support our initial results: a small positive slope was observed (in contrast to the previous analysis) that did not reach statistical significance in either of the linear regression methods ($p = 0.3832$ and $p = 0.831$) (Figure 2 (center)). Moreover, in a combined analysis of both SALS sets (859 subjects), neither linear regression model reached statistical significance ($p = 0.1436$ and $p = 0.2282$, respectively) (Figure 2 (right)).

Discussion

Our study aimed to clarify the role of the HTT CAG repeat in SALS. Recent studies reported an association between the ataxin-2 CAG repeat and ALS. These findings raised the possibility that such CAG expansions could contribute to ALS in a more general way, possibly by having an impact on the cellular protein degradation systems (23,24), and therefore raised the question of whether this effect is specific of ataxin-2 or, alternatively, whether other polyQ disease proteins, such as huntingtin, also contribute to ALS neurodegeneration. A potential relationship between the mechanisms underlying HD and ALS was also suggested by case reports of motor neuron disease associated with HD. Although association of chorea with motor neuron disease is rare, HD patients with (concurrent) motor neuron signs and clear pathological evidence of ALS have been found (25,26). This co-occurrence may suggest that...
existence of two separate disease processes or the result of progression of HD pathology to motor neurons. The temporal coincidence of motor neuron and striatal involvement in the patient described by Sadeghian et al. argues for early HD manifesting in both neuronal systems rather than chance occurrence of independent disease processes (26). Furthermore, chorea has also been sporadically described in patients with ALS without HD, suggesting that a prolonged course of ALS may result in an impact of degeneration on extrapyramidal structures, with later development of hyperkinetic movements (27,28). Additionally, optineurin, occasionally found in cytoplasmic granules of HD patients’ neurons and known to interact with huntingtin, has recently been found to be mutated in FALS cases (11). All these data supported the potential for a role for huntingtin in ALS pathogenesis.

Our results show that the HTT CAG repeat has no effect on susceptibility to ALS or to its phenotypic manifestations, including age at onset, site of onset and survival after onset. Although a small negative trend was observed between HTT CAG repeat size and ALS age at onset, it did not reach statistical significance. This result does not support the role of HTT CAG as a genetic modifier of ALS clinical presentation, either modifying age at onset or impacting on the ALS clinical subtype (no HD repeat effect was found either in bulbar or spinal onset cases). Similarly, although a small negative effect (~1 month per additional repeat) of the HTT CAG repeat on ALS survival was suggested initially, further investigation did not support this contention. It should be taken into consideration that this analysis was performed only on deceased ALS cases and that the samples were derived from tertiary referral centers, which in combination could have introduced some bias to the survival analysis. If the HD CAG repeat is indeed an ALS survival risk factor, we would expect that an increase on the repeat size would lead to a worse survival and therefore very fast progressive patients could have been missed in this study, diluting the effect of a possible increased risk. To better answer this question, a population based sample study with complete information on survival status should be used.

Recent data corroborate our findings and reinforce a specific role for ataxin-2 in ALS rather than the general polyQ hypothesis (29). The role of ataxin-2 CAG repeat length in neurodegeneration was recently also explored in patients with frontotemporal lobar degeneration, progressive supranuclear palsy, Alzheimer’s and Parkinson’s disease, and although not showing an impact in all these disorders, ataxin-2 repeats may predispose to certain types of neurodegeneration, in particular to progressive supranuclear palsy neurodegeneration (30). Our data reveal no significant association between the HTT CAG repeat and ALS, which makes a general polyQ impact on ALS pathogenesis improbable, suggesting instead different neuronal degeneration processes in HD and ALS.

Acknowledgements

The authors wish to thank Lucie Bruijn of the ALS Association, Seung Kwaek, David Howland and Jamshid Arjomand of CHDI Foundation for enabling the study.

EMR is the recipient of a scholarship from Fundação para a Ciência e a Tecnologia. Grants were received from the National Institutes of Health NINDS Huntington’s Disease Center Without Walls (JFG, MEM) and the CHDI Foundation, Inc. (JFG, MEM). Generous support was provided by the ALS Therapy Alliance, Project ALS, the Angel Fund, the Pierre L. de Bourgknecht ALS Research Foundation, the Al-Athel ALS Research Foundation, the ALS Family Charitable Foundation and the National Institute of Neurological Disorders and Stroke (RHB). JEL acknowledges grant support from NIH/NINDS. NT and VS have been supported by a Francesco Caleffi donation and acknowledge grant support from ArtiSAL and the Italian Ministry of Health, Ricerca Finalizzata, 2009.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Results – Article 8

HD CAG repeat length in ALS patients


Article 9: Prevalence of Huntington’s disease CAG trinucleotide repeat alleles in bipolar disease patients.

Ramos EM, Gillis T, Mysore JS, Lee JM, Alonso I, Gusella JF, Smoller JW, Sklar P, MacDonald ME, Perlis R.

[submitted]
Prevalence of Huntington’s disease gene CAG trinucleotide repeat alleles in bipolar disorder patients

Eliana Marisa Ramos 1,2, Tammy Gillis 1, Jayalakshmi S. Mysore 1, Jong-Min Lee 1, Isabel Alonso 2, James F. Gusella 1, Jordan W. Smoller 1, Pamela Sklar 3, Marcy E. MacDonald 1, Roy H. Perlis 1*

1 Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 2 UnIGENe, IBMC – Institute for Molecular and Cell Biology, University of Porto, Porto, Portugal; 3 Division of Psychiatric Genomics, Department of Psychiatry, Mount Sinai School of Medicine, New York, NY, USA

*Corresponding author Dr. Roy H. Perlis, M.D., Center for Human Genetic Research, Massachusetts General Hospital, Simches Research Building, 185 Cambridge Street, Boston MA 02114
Telephone: 617 726 7426
Email: rperlis@partners.org

Running title HD CAG repeat in bipolar disorder

Abstract Objectives: Huntington’s disease is a neurodegenerative disorder characterized by motor, cognitive and psychiatric symptoms that are caused by HTT CAG trinucleotide repeat alleles of 35 or more units. A greater than expected prevalence of incompletely penetrant HTT CAG repeat alleles among individuals diagnosed with major depressive disorder has raised the possibility that related psychiatric disorders could also be associated with Huntington’s disease. Methods: Here we have assessed the distribution of HTT CAG repeat alleles in bipolar disorder, a mood disorder that typically includes depressive as well as manic or hypomanic episodes. HTT CAG allele sizes from 2,229 Caucasian individuals diagnosed with DSM-IV bipolar disorder were compared to allele sizes in 1,828 control individuals. Results: We found that HTT CAG repeat alleles greater than 35 units were observed in only one of 4,458 chromosomes from individuals with bipolar disorder, compared with three of 3,656 chromosomes from control subjects.
Conclusions: These findings suggest, therefore, that the clinical diagnosis of bipolar disorder is unlikely to be confounded by Huntington’s disease.

Keywords Huntington’s disease; bipolar disorder; trinucleotide repeat; depression; neurodegenerative disease; polyglutamine expansion

Introduction Huntington’s disease (HD) is a dominantly inherited progressive neurodegenerative disorder caused by the expansion of an unstable polymorphic CAG repeat in the HTT gene (chromosome 4p16.3), resulting in an expanded polyglutamine tract in the huntingtin protein (1). Alleles with fewer than 35 CAGs are considered to produce no overt clinical symptoms of HD and can be grouped in two categories: stable normal alleles (with less than 27 CAGs), and high-normal alleles, which have in some cases been reported to expand (27-35 CAGs) upon transmission. A third class of alleles between 35-39 repeats shows incomplete HD penetrance while alleles with 40 or more repeats are fully penetrant and lead to the onset of HD symptoms. HD is mainly characterized by motor, cognitive and behavioral manifestations. The psychiatric symptoms that become manifest in many, but not all, HD patients, tend to be depression and apathy but mania has also be observed; elevated or irritable mood, overactivity, decreased need for sleep, hypersexuality, impulsiveness and grandiosity, and severe cases may have delusions and hallucinations (2). Indeed, HD patients who alternate between continued periods of depression and mania, interspersed with periods of normal mood, have been misdiagnosed with bipolar disorder. This major affective mood disorder, which has a prevalence of up to 3% (3) in the general population, features episodic periods of abnormally elevated mood states (mania, hypomania, or mixed states) and typically depressed states as well. Furthermore, depending upon age of onset and severity of clinical symptoms, some bipolar patients may initially be diagnosed and treated for major depression (4).

Recently, we found that HD-associated CAG repeat alleles were overrepresented among individuals diagnosed with major depression, estimating that 3.3 in 1,000 individuals diagnosed with major depression carry an expanded HTT CAG allele (5). These findings highlighted the importance of considering HD in the differential diagnosis of major depressive disorder. They also strongly implied that HTT CAG allele sizes, in the reduced- and fully- penetrant ranges, may act to sensitize individuals to the effects of genetic or environmental factors in the general population that contribute to the manifestation of clinical depression both in HD patients and in individuals diagnosed with major depressive disorder. Here we have evaluated whether there may, or may not, be an overlap in the
underlying genetic architecture of the biology of HD and bipolar disorder. We determined how often individuals diagnosed with bipolar disorder carry HTT CAG alleles that are associated with HD, by genotyping the HTT CAG repeat in 2,529 individuals with a DSM-IV diagnosis of bipolar disorder.

Material and Methods

Bipolar disorder cohort. The patient cohort comprised a total of 2,529 individuals diagnosed with bipolar disorder. From these, 1,277 individuals were diagnosed with bipolar I (mainly characterized by episodes of mania) while 760 had bipolar II (alternated episodes of hypomania and major depression). In this cohort, 57.1% (1,443) were women and 88.6% had Caucasian, 6.6% African and 2.2% Asian (self-reported) ancestry.

Control cohort 1. This cohort comprised the normal chromosomes 4 of a large cohort of 4,007 HD individuals, utilized previously in studies that assessed the HTT CAG repeat allele distribution in major depression and amyotrophic lateral sclerosis patients (5, 6). These chromosomes, with the shorter non-expanded HTT CAG repeat alleles (less than 35 repeats), should reflect the distribution of HTT CAG alleles in the general population as HD is a true dominant disorder (7) that is relatively rare (estimated at 1 in 7-10,000 Caucasian individuals).

Control cohort 2. A cohort of 1,288 psychiatrically-screened controls (with no diagnosis of bipolar disorder or schizophrenia) and 547 DNA samples derived from anonymous cord blood donors, were used as control group 2 (8). In this control group all individuals were Caucasian (self-reported) and 49.0% (899) were women.

Genotyping. HTT CAG repeat allele sizes were determined by a polymerase chain reaction (PCR) amplification assay, using fluorescently labelled primers, as previously described (9). The PCR products were run on an ABI PRISM 3730XL automated DNA Sequencer (Applied Biosystems, Foster City, CA) and analysed using GeneMapper version 3.7 software. A set of genomic DNA standards for which HTT CAG repeat number had been confirmed by DNA sequencing, was used to provide CAG allele size standards.

Statistical analysis. Fisher’s exact test was used to compare HTT CAG repeat frequency between the bipolar disorder individuals and each of the two different comparison control cohorts. We compared the repeat frequency of normal (with less than 27 CAGs) and high-normal alleles (27-35 CAGs), as well as stable normal alleles and alleles with a tendency to expand (>26 CAGs). The HTT CAG distribution was also analyzed using the nonparametric Mann-Whitney U test, since the samples did not display a normal
distribution. All the statistical analyses were performed using PASW Statistics 18 (SPSS Inc., Chicago, IL).

**Results** In the group of individuals with bipolar disorder, the HTT CAG repeat range varied between nine to 36 repeats, with two out of 5,032 chromosomes having 35 CAG repeats (end of high-normal range) and only one chromosome carried a reduced-penetrance allele with 36 CAG repeats. The individual carrying a 36 CAG repeat was diagnosed with bipolar II, while the two individuals with 35 CAG repeats were bipolar I and II cases. In order to characterize the prevalence of HTT CAG repeats in bipolar disorder, we used as comparison group for the CAG repeat frequency and distribution, 4,007 non-HD control chromosomes (control cohort 1) that was also the subject of a prior analysis of major depressive disorder (5). As previously reported, the non-HD control chromosomes HD CAG range varied between nine and 34 HTT CAG repeats (5, 6) with none of these chromosomes having 35 or more CAG repeats. The allele frequencies in the bipolar disorder individuals were not significantly different from the non-HD control chromosomes (Fisher’s exact test: normal versus high-normal alleles, \( p=0.86 \) and normal versus >26 CAGs, \( p=0.99 \)) (Table 1). However, comparison of the distributions of the allele sizes, assessed using the nonparametric Mann-Whitney U test, revealed a statistically significant difference between the non-HD control and bipolar disorder chromosomes (Mann-Whitney U test, \( p= 0.036 \)) (Figure 1).

Since the cohort 1 control group, comprising the non-HD chromosomes from HD patients (defined as 34 or fewer repeats) may underestimate the true frequency of alleles of 35 or more CAG repeats in the general population, we genotyped a second large control cohort, comprising 1,288 psychiatrically-screened controls and 547 cord blood DNA samples. In this control cohort 2, the HTT CAG repeat ranged from nine to 37 CAG repeats, with three alleles having high-end normal alleles (35 CAGs) and three alleles in the reduced penetrance range (one 36 CAG and two 37 CAG repeat alleles) (Table 1). When comparing allele frequencies of bipolar disorder subjects with this set of control individuals, there was a significant difference when comparing both normal versus high normal (Fisher’s exact test, \( p=0.011 \)) and normal versus >26 HD CAG alleles (Fisher’s exact test, \( p=0.08 \)) alleles.

To evaluate whether the apparent under-representation of high-end normal and HD-associated alleles in the bipolar disorder cases might reflect population stratification due to ancestry differences between the disease cases and the control samples (all self-reported Caucasian), we excluded from the bipolar disorder cohort all samples of non-Caucasian origin. This yielded a total of 4,458 chromosomes from Caucasian individuals diagnosed
with bipolar disorder. In this final set, the HTT CAG range and frequency of high-end normal and reduced penetrance alleles remained similar (sizes from nine to 36 repeats, with a single reduced-penetrance allele). After controlling for ancestry, there was a significant difference between the allele frequency of normal versus high-normal and normal versus >26 CAG repeat alleles (Fisher’s exact test, \( p = 0.040 \) and \( p = 0.030 \) respectively). However, the Mann-Whitney U test revealed no significant difference in the distributions of the HTT CAG allele sizes between the control and bipolar disorder Caucasian chromosomes (Mann-Whitney U test, \( p = 0.216 \)) (Figure 2).

**Discussion** The genetic architecture of major psychiatric disorders is complex. One genetic contributor highlighted by psychiatric symptoms in a proportion of HD patients, may be the HTT CAG repeat, at least in the range of allele sizes reported to be associated with a clinical diagnosis of HD. We speculate that these HTT CAG repeat alleles, which encode a functionally important run of polyglutamines in the huntingtin protein (10), may sensitize certain individuals in the general population to genetic and environmental factors that increase the risk of developing particular psychiatric symptoms. For example, overrepresentation of HD-associated CAG alleles appears to further increase the risk of developing depression, such that HD should be considered in the clinical diagnosis of major depression (5). Our current study aimed to determine the prevalence of HD-associated HTT CAG repeat alleles in a clinical population of individuals diagnosed with bipolar disorder, to determine whether these alleles may contribute to the genetic architecture of this major affective mood disorder, which, in addition to episodes of mania (seen in some HD patients) features depressive episodes.

Our results disclosed that 1 in 2,229 individuals, or about 0.45 in every 1,000 individuals, diagnosed with bipolar disorder carry an expanded HTT CAG allele in the range that has been associated with development of clinical symptoms of HD, whereas our results demonstrate that approximately 1.64 in every 1,000 individuals from the general population carry an expanded HTT CAG allele, a value similar to that reported in the literature (~2.26 in 1,000 in the Portuguese population) (11). Thus, in contrast to major depressive disorder, HD seems unlikely to be a significant confound in the differential clinical diagnosis of bipolar disorder.

One caveat to this interpretation, which is relevant to all case-control studies, is the composition of the control comparison group, which should be representative of the general population and provide the best estimate of the expected disease prevalence. In previous studies, we have used as a comparison group the non-HD chromosomes from a large collection of individuals diagnosed with HD. We believe this group provides a good
representation of the $HTT$ CAG allele distribution in the general population because HD is a truly dominant disorder (7) that is relatively rare (~1 in 7-10,000). Therefore, we would expect the non-HD chromosome to be drawn from the general population and would not expect to find any alleles higher than 35 CAGs in a random set of 5,000 individuals. Nevertheless, this selection criterion excludes reduced-penetrant as well as fully-penetrant HD alleles, which occur at some level in the general population. To eliminate this potential bias, we utilized a second comparison group that included individuals screened to exclude bipolar disorder and schizophrenia. The two main advantages of this group are that both $HTT$ CAG alleles in all individuals contribute to the allele frequency and these provide an optimal match with the bipolar disorder cohort with respect to their geographic and genetic origin, thereby avoiding, or at least decreasing, the contribution of the effects of population stratification in the analysis. This is important because there are significant geographic differences in the prevalence and distribution of the $HTT$ CAG repeat, related to different ancestral $HTT$ haplotypes (12, 13). The modal $HTT$ CAG repeat length in European and East Asian populations is 17 while African populations have a broader $HTT$ CAG distribution, with most of the alleles having 15 CAG repeats (14, 15). However, despite the apparent advantages, the criteria used to select this second cohort may themselves introduce different biases into the analysis. About 2/3 of the individuals included in this non-psychiatric control group did not have a DSM-IV diagnosis of bipolar disorder and schizophrenia, while 1/3 control group represented anonymous cord blood samples. While the latter may represent general population, the former is not a true random sample. Furthermore, none of these controls were screened for major depression disorder. The lifetime prevalence of major depression is roughly 17% (16), therefore we cannot exclude the possibility that the reduced penetrance HD alleles observed in this cohort could be associated with individuals with major depression; previously reported to have an enrichment of these alleles (5). In addition, the level of neurologic examination and collection of family neurological history may not have been sufficient to exclude a diagnosis of HD.

Our results do not support an obvious association between $HTT$ and bipolar disorder. This stands in contrast to the overrepresentation in individuals diagnosed with major depressive disorder. This apparent difference, thereby, implies that in some respects bipolar disorder and major depression may have distinct genetic architectures; with only the latter being sensitive to the effects of the $HTT$ CAG repeat. Further studies examining the prevalence of the $HTT$ CAG repeat in other psychiatric disorders, such as schizophrenia, among others, may help uncover a possible common genetic mechanism
in a subset of disorders that is sensitive to the effects of the CAG repeat that is the root cause of HD.

Acknowledgments The authors wish to thank Drs. Seung Kwak, Jamshid Arajomond and Ignacio Muñoz-Sanjuan of the CHDI Foundation, Inc.

Funding Sources EMR is the recipient of a scholarship from Fundação para a Ciência e a Tecnologia (SFRH/BD/44335/2008). Grants from the National Institutes of Health NINDS NS16367 Huntington's Disease Center Without Walls (JFG, MEM), NS32765 (MEM) and the CHDI Foundation, Inc. (JFG, MEM). Dr. Perlis is supported by R01MH086026 from the National Institute of Mental Health.

Disclosure of Interests Dr. Perlis has received consulting fees or served on scientific advisory boards for Proteus Biomedical, Healthrageous, Pamlab, Genomind, and RIDventures, research grant support from Proteus Biomedical, and royalties from Concordant Rater Systems.

References


Tables and Figures

Table 1. *HTT* CAG repeat alleles in bipolar disorder

<table>
<thead>
<tr>
<th></th>
<th>Normal HD allele (&lt;27 CAGs)</th>
<th>High normal HD allele (27-35 CAGs)</th>
<th>At risk HD allele (&gt;35 CAGs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bipolar Disorder chromosomes</td>
<td>4890 (97.18%)</td>
<td>141 (2.80%)</td>
<td>1 (0.02%)</td>
</tr>
<tr>
<td>Caucasian</td>
<td>4326 (97.04%)</td>
<td>131 (2.94%)</td>
<td>1 (0.02%)</td>
</tr>
<tr>
<td>Control chromosomes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohort 1</td>
<td>3869 (96.56%)</td>
<td>138 (3.44%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Cohort 2</td>
<td>3515 (96.14%)</td>
<td>138 (3.77%)</td>
<td>3 (0.08%)</td>
</tr>
</tbody>
</table>

Figure 1. *HTT* CAG repeat length in 4,007 non-HD control chromosomes (cohort 1) and 5,032 bipolar disorder chromosomes. Insert shows a zoomed view of high-normal and reduced penetrance HD CAG alleles (27 to 38 CAG repeats).
Results – Article 9

Figure 2. *HTT* CAG repeat length in 3,656 control chromosomes (cohort 2) [from 1,288 psychiatrically-screened and 547 cord blood control samples] and 4,458 bipolar disorder chromosomes, where all individuals have Caucasian origin. Insert shows a zoomed view of high-normal and reduced penetrance HD CAG alleles (27 to 38 CAG repeats).
Validation in Huntington's Disease Experimental Models

Hdh knock-in mice, expressing the full-length HD mutant protein, replicate precisely the CAG repeat mutation in the mouse homolog and is therefore one of the best model systems to help define the mechanism of action of the genetic variation that has a modifier effect. Furthermore, since modifier genes require the presence of the conditioning mutation, chromosome substitution strains (CSS) are a valuable cross-species functional approach for experimental validation and study of human modifier loci from unbiased human genetic studies, which typically are large making it hard for unbiased gene-by-gene analyses.

Seeing that initial linkage analyses have identified a large linkage region on the human 6q23-24 region, we have now evaluated the potential functional effect of variants in this region in the onset and severity of early effects of the CAG repeat mutation. For that, we bred C57BL6/J Hdh CAG repeat knock-in mice with CSS10 mice to generate an F1 cross of HdhQ111/+B6.AJ10 mice, that carries one chromosome 10 (chr10) that is A/J but the rest of the genome is C57BL6/J. By comparing a relatively small number of HdhQ111/+B6 and HdhQ111/+B6.AJ10 F1 mice, we demonstrated that the presence of a single A/J chr10 (corresponding to the human chromosome 6) could mildly alter early, dominant and CAG length-dependent HD phenotypes, such as HTT somatic liver instability and the appearance of NIs of mutant huntingtin, while it did not seem to affect other phenotypes, such as body weight and levels of DARPP-32 in medium-spiny neurons [Article 10].
Article 10: Chromosome substitution mouse strain approach to functional assessment of the 6q23-24 Huntington’s disease modifier locus.


[in preparation]
Chromosome substitution mouse strain approach to functional assessment of Huntington’s disease modifier locus (6q23-24)

Eliana Marisa Ramos, Jolene R. Guide, Jason St. Claire, Tammy Gillis, Jayalakshmi S. Mysore, Marina Kovalenko, Jorge Sequeiros, Vanessa Wheeler, Isabel Alonso, Marcy E. MacDonald

From the Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA (E.M.R., J.R.G., J.S.C., T.G., J.S.M., M.K., V.W., M.E.M.); UnIGENe, IBMC – Institute for Molecular and Cell Biology, University of Porto, Porto, Portugal (E.M.R., J.S., I.A.); CGPP, IBMC – Institute for Molecular and Cell Biology, University of Porto, Porto, Portugal (J.S., I.A.); ICBAS – Instituto de Ciências Biomédicas Abel Salazar, University of Porto, Porto, Portugal (J.S., I.A.)

Corresponding Author Marcy E. MacDonald, Center for Human Genetic Research, Massachusetts General Hospital, Simches Research Building, Room 5414, 185 Cambridge Street, Boston, MA 02114, USA
Telephone: +1(617)-726-5089
Fax: +1(617)-726-5735
Email: macdonam@helix.mgh.harvard.edu

Abstract Huntington’s disease (HD) is an autosomal dominant neurodegenerative disorder characterized by motor, cognitive and psychiatric manifestations. Even though HD is caused by the expansion of an unstable polymorphic CAG repeat in the HTT gene, the disease expression and its timing are clearly modifiable by other factors. In order to evaluate the effect of a previously HD associated linkage region at the large 6q23-24 interval, we used a novel human-mouse cross-species approach. This unbiased strategy employs C57BL/6J (B6J) Hdh\textsuperscript{Q111} knock-in mice, which precisely replicate the HD mutation, and the C57BL/6J-chr10\textsuperscript{AJ}/NaJ chromosome substitution strain (CSS10), in which chromosome 10 (chr10), in synteny with the human 6q23-24 region, is derived from the A/J (AJ) strain while the other chromosomes are derived from the B6J recipient strain. We crossed wild type CSS10 and B6J mice with Hdh\textsuperscript{Q111/+} knock-in mice to ask whether
dominantly acting chr10 AJ-B6J variants were sufficient to alter a panel of early, dominant and CAG length-dependent phenotypes. Our results revealed that the presence of a single copy of an AJ chromosome had significant effects on the rate of body weight, regardless of the HD mutation, and in HdhQ111 had mildly effects on the levels and patterns of somatic CAG instability in the liver and the appearance of mutant neuronal nuclear inclusions, as well as DARPP-32 levels in medium-spiny neurons. These findings suggest a potential mild chr10-wide effect on some of the mechanisms involved in the HD pathogenesis, and encourage a larger study with CSS10 and sub-strains. We believe that the use of a CSS panel to screen for modifiers in HD knock-in mice might help guide the unbiased human genome scan studies and speed up the discovery of candidate genes that delay HD phenotypes and could therefore have a tremendous therapeutic impact.

**Keywords** Huntington’s disease; knock-in mice; consomic mouse; genetic modifiers

**Introduction** Huntington’s disease (HD) is a dominantly inherited progressive neurodegenerative disorder, usually of adult onset, characterized by involuntary choreic movements, cognitive impairment and behavioral changes. It is caused by the expansion of an unstable polymorphic CAG repeat in the HTT gene [1]. The expanded CAG repeat is the major determinant of age-at-onset (AO) of motor symptoms, such that the longer the repeat the earlier the onset. It explains about 50-70% of the variance in motor AO [2-4], while the remainder (residual AO) is highly heritable strongly implying the existence of genetic factors that modulate the rate of the pathogenic process that leads to onset of symptoms [5-7].

In the never ceasing search for modifier genes that could explain the residual onset of HD symptoms, a whole-genome scan in affected HD sibling pairs, the HD-MAPS study, has provided strong evidence for a quantitative trait loci (QTL) at the human chromosomal region 6q23-24 modifying neurological onset [8, 9]. This 6q23-24 interval contains at least 128 known and predicted genes [8], making it hard to test the potential association of each gene with HD. Therefore, and in order to screen and validate this region for genetic modifiers of HD onset, we decided to take advantage of a recent mouse model approach used to identify and validate QTLs, the chromosome substitution mouse strains (CSSs). These mice, commonly referred as CSSi, carry both copies of a chromosome i from a donor strain, e.g. A/J (AJ), while all other chromosomes from the host strain, e.g. C57BL/6J (B6J) are intact and homozygous [10]. By testing mice from a CSS strain for a phenotype of interest it can be immediately inferred that phenotypic differences between the CSS and the host strain implies that at least one QTL resides on the substituted
chromosome. The same principle can be used to identify regions carrying potential genetic modifiers of a particular phenotype. The cross between CSSs mice and mice carrying a mutation known to cause a human disease, such as the Hdh\(^{Q111}\) knock-in mice [11, 12] that replicates the HD genetic mutation, provides an ideal model to study genetic factors in the substituted chromosome that may influence disease phenotypes, such as the onset of HD symptoms. The Hdh\(^{Q111}\) knock-in mice exhibit accurate expression of mutant huntingtin and display early presymptomatic phenotypes that are dominant over the normal protein, CAG repeat length-dependent and occur with strong selectivity in medium spiny neurons (MSNs) [13, 14], key features of the human HD mechanism. Theoretically, a difference in any of these phenotypes between the F1 progeny and parental knock-in mice implies that at least one or more genetic modifiers of HD reside on the substituted chromosome.

In this study, we sought to explore and validate the previous association of the human 6q23-24 region with HD onset. Since the human chromosome 6 is in syntenic with the mouse chromosome 10 (chr10), we used a cross between CSS10 mice and our Hdh\(^{Q111}\) knock-in mice to generate progeny with one copy of an AJ chr10 and heterozygous for the HD mutation in an otherwise complete B6J background. By comparing these F1 mice with our Hdh\(^{Q111}\) knock-in mice we are testing if chr10 AJ-B6J genetic variants are sufficient to alter early, dominant and CAG length-dependent phenotypes, such as Htt CAG somatic instability [12, 15, 16], intranuclear inclusions of mutant huntingtin [14] and expression of dopamine- and cAMP-regulated phosphoprotein, 32 kDa (DARPP-32).

**Material and Methods**

**Mice.** In this study we obtained from The Jackson Laboratories (Jackson Laboratories, Bar Harbor, ME) C57BL/6J-Chr10\(^{AJ}/NaJ\) (CSS10) mice that carry both chr10 from strain AJ on an otherwise B6J background [10], and wild type B6J mice. Hdh\(^{Q111}\) knock-in mice used in this study were on a B6J background [11, 12], and the actual Htt repeat size of these mice was 139 CAG repeats. As shown in Figure 1, crosses between Hdh\(^{Q111}\) knock-in and wild type (CSS10 and B6J) mice were made in one direction – Hdh\(^{Q111}\) male x B6J and CSS10 females – in order to control for possible parental effects of the Htt mutant allele. Briefly, male heterozygous Hdh\(^{Q111}\) knock-in mice were crossed with CSS10 female mice in order to generate wild type Hdh\(^{+/-}\).C57BL/6J.Chr10\(^{AJ}/NaJ\) (Hdh\(^{+/-}\)B6J.AJ10) and mutant Hdh\(^{Q111+/+}\).C57BL/6J.Chr10\(^{AJ}/NaJ\) (Hdh\(^{Q111+/+}\)B6J.AJ10) mice that carried one chr10 from strain AJ on an otherwise B6J background. Subsequently, the same male heterozygous Hdh\(^{Q111}\) knock-in mice were crossed with B6J female mice in order to generate wild type Hdh\(^{+/-}\).C57BL/6J (Hdh\(^{+/-}\)B6J) and mutant Hdh\(^{Q111+/+}\).C57BL/6J (Hdh\(^{Q111+/+}\)B6J) mice on a
Results – Article 10

Complete B6J background. Animal experiments were performed to minimize pain and discomfort, under an approved protocol of the Massachusetts General Hospital Subcommittee on Research Animal Care (SRAC, protocol number: 2005N000112).

Experimental design. From the F1 offspring mice (n=119), we selected two different sets of mice for somatic instability (set 1) and immunohistochemistry assays (set 2), with groups consisting of both female and male, Htt mutant and wild type mice from both B6J.AJ10 and B6J crosses (Supplementary Table 1). Set 1 (n=20) consisted of ten 5 months old Htt mutant mice per genotype (Hdh\textsuperscript{Q111/+}B6J.AJ10 and Hdh\textsuperscript{Q111/+}B6J) matched by gender (5 female and 5 male per genotype) and Htt CAG repeat size. Set 2 (n=40) consisted of ten 5 months old mice per genotype (Hdh\textsuperscript{Q111/+}B6J.AJ10, Hdh\textsuperscript{Q111/+}B6J, Hdh\textsuperscript{+/+}B6J.AJ10 and Hdh\textsuperscript{+/+}B6J) matched by gender (5 females and 5 males per genotype) and Htt CAG repeat size. All Htt mutant mice used in this study were heterozygous for the Htt CAG mutation. As we could not generate data for all the animals in all assays, the number of mice included into the analyses of each test is given in the legends.

Body weight. F1 offspring mice were fed Prolab Isopro RMH 3000 (PMI Nutrition International, Brentwood, MO) ad libitum and weighted at weekly intervals from the time of weaning (3 weeks) until 5 months (~24 weeks). The following body weight traits were analyzed: initial (IW: weight in grams at ~25 days of age), medium (MW: weight in grams at ~98 days of age) and final weight (FW: weight at ~169 days of age). We also calculated the mean weight gain/day in grams for the first (EWG) and last ~72 days (FWG) and the weight gain per day in grams/day (WG), as previously described [17].

Genotyping and analysis of somatic instability. Genomic DNA was isolated from fresh-frozen collected tissues (tail, striatum and liver) using DNeasy Blood & Tissue kit (Qiagen, Valencia, CA). Genotyping of the Hdh\textsuperscript{Q111} knock-in allele was carried out using a previously established polymerase chain reaction (PCR) amplification assay, with fluorescently labelled primers [18]. The size of the PCR products was then determined using the ABI PRISM 3730x/ automated DNA Sequencer (Applied Biosystems, Foster City, CA) and GeneMapper version 3.7 software. All runs included the same control DNAs of known Htt CAG repeat size. Somatic instability was quantified from the GeneMapper traces as described previously [15]. Briefly, the main allele was identified as the highest peak in each analysis and peaks with height less than 20% of the main allele were excluded. The peak height of each peak was divided by the sum of the heights of all signal peaks and then multiplied by the CAG change relative to the main allele. These values were summed to generate an instability index. To qualitatively assess the different patterns of repeat instability in liver and striatum, the distance (in CAG repeats) between
the two modes and the distance (in CAG repeats) between the constitutive repeat mode and the longest repeat was calculated as previously described [16].

**Immunohistochemistry.** Coronal sections of paraformaldehyde-perfused and post-fixed, gelatin-embedded hemisphere brains were performed at Neurosciences Associates (NSALabs, Knoxville, TN) using MultiBrain Technology. With this technology, all mouse brain hemispheres were embedded in one single block that was freeze-sectioned at 35μ in the coronal plane (throughout the striatum) with a sliding microtome, resulting in free-floating sections. Designated sections across the striatum were stained with either EM48 and thionine counterstain or DARPP-32 at 210μ intervals. Bright field microscopy was performed with an Olympus BX51 microscope equipped with a Qcolor5 Olympus camera and Qcapture image acquisition software. Images that were to be quantified and compared were taken with the same exposure times. For the EM48 sections, three micrographs (40x objective) were taken from the dorsal medial striatum in three consecutive sections from each mouse. The number of nuclear huntingtin inclusions was quantified in three striatal areas per mouse and normalized to the number of EM48-positive cells. For the DARPP-32 sections, micrographs (4x objective) in three consecutive sections were taken from each mouse and DARPP-32 intensity was quantified using ImageJ image analysis software [19].

**Statistical analyses.** Pairwise comparisons were determined using an unpaired $t$-test or a nonparametric unpaired Mann Whitney test (for data that did not display a normal distribution). All statistical analyses were performed using GraphPad Prism 6.00 (GraphPad Software, La Jolla, CA).

**Results**

**Genetic variance between B6J and AJ strains at mouse chr10.** A whole-genome scan has provided strong evidence for a QTL, between markers D6S1009 and D6S2436, at the human 6q23-24 modifying neurological onset in HD [8, 9]. This region (~16Mbp) includes at least 80 different human genes that corresponded to 72 homologue genes in the mouse, as annotated in the Ensembl Genome Browser [20]. The human 6q23-24 is in synteny with part of mouse chr10 (Figure 2A), comprising 71 of the homologue genes. According to the Mouse Genome Informatics database [21], B6J and AJ mice have about 16,000 allelic differences at chr10 (~131Mbp): 106 non-synonymous, 205 synonymous, ~14,600 intronic, 576 locus-region (in an untranscribed region flanking a gene: 2Kb 5’ or 0.5Kb 3’), 316 mRNA-UTR (in the transcript but not in the coding region interval) and 3 splice-site (in the first or two last bases of an intron) variants (Figure 2B). From these variants, nineteen coding non-synonymous SNPs, most probable to have functional
Results – Article 10

CAG Instability of Hdh\textsuperscript{Q111} knock-in alleles. Previous studies have shown that in Hdh\textsuperscript{Q111} knock-in mice intergenerational CAG repeat length changes with a paternal expansion bias [12, 22]. In our study, the Htt mutant allele was transmitted from Hdh\textsuperscript{Q111} knock-in mice with the same background (B6J) and original CAG repeat length (139 CAG repeats). Therefore, as expected, we did observe a paternal expansion bias (74% in crosses of Hdh\textsuperscript{Q111} mice with B6J and 82% with CSS10 mice), but no significant repeat length difference in the transmission of the Htt 139 CAG knock-in allele to Hdh\textsuperscript{Q111}/\textsuperscript{B6J} (mean ± SD = 3.04±3.18) and Hdh\textsuperscript{Q111}/\textsuperscript{B6J.AJ10} mice (mean ± SD = 2.25±2.77). Figure 3 shows the actual CAG repeat size of the Hdh\textsuperscript{Q111} knock-in allele in both Hdh\textsuperscript{Q111}/\textsuperscript{B6J} and Hdh\textsuperscript{Q111}/\textsuperscript{B6J.AJ10} progeny used in this study. Hdh\textsuperscript{Q111} knock-in mice also exhibit CAG length, age dependent and tissue specific somatic instability, with significant accumulation of expansions in striatum and liver that becomes apparent by 5 months of age [12, 15, 16].

In order to evaluate the potential effect of AJ-B6J genetic variants at chr10 in somatic instability, we extracted genomic DNA from one stable (tail) and two unstable (liver and striatum) tissues from Hdh\textsuperscript{Q111}/\textsuperscript{B6J} and Hdh\textsuperscript{Q111}/\textsuperscript{B6J.AJ10} mice at 5 months of age. As shown in Figure 4A, the Htt CAG repeats in tail were very stable in contrast to striatum and liver that showed significant levels of instability, irrespective of chr10 background. However, we did find one female Hdh\textsuperscript{Q111}/\textsuperscript{B6J} mouse (with 139 CAG repeats) with an unstable CAG repeat in the tail (Figure 4B), where we could observe a distinct population of unstable repeats that were evident in the tail already at 3 weeks of age, and in the striatum and liver since at least 5 months of age. We excluded this mouse from our somatic instability analysis. Next we quantified the levels of repeat instability for each tissue using a conservative threshold factor (20% of the highest peak) as it is more resistant to the amplification variation that we observed in our study. As shown in Figure 4C, the highest instability indices were observed in the liver (+11.95±0.81 for Hdh\textsuperscript{Q111}/\textsuperscript{B6J} and +12.71±1.03 for Hdh\textsuperscript{Q111}/\textsuperscript{B6J.AJ10}) followed by the striatum (+8.15±1.51 for Hdh\textsuperscript{Q111}/\textsuperscript{B6J} and +8.40±1.27 for Hdh\textsuperscript{Q111}/\textsuperscript{B6J.AJ10}). We did not find any significant difference for the instability indices, nor for the expansion and contraction indices (data not shown) when comparing our two groups of Hdh\textsuperscript{Q111} mice. As previously described [16], the patterns of repeat instability differed between liver and striatum: while in the liver we could observe a distinct population of unstable CAG repeats, in the striatum these repeats were more broadly distributed. To qualitatively characterize these different patterns of repeat instability, two additional measurements were made: the distance to the longest repeat

impact, are located in twelve homologue genes of human genes located at 6q23-24: Mthfd1l, 1700052N19Rik, Katna1, Sash1, Samd5, Epm2a, Utrn, Stx11, Hivep2, Nhs1l, Tnfaip3, Ifngr1.
Results – Article 10

Results – Article 10

There was no significant difference in the patterns of repeat instability when comparing Hdh\(^{Q111/4}\)B6J and Hdh\(^{Q111/4}\)B6J.AJ10, however, as for the instability index, we did observe a numerically larger distribution of the unstable repeats on the liver of Hdh\(^{Q111/4}\)B6J.AJ10 mice.

**Body Weight.** In patients with HD, weight loss is often observed in the course of the disease [23-27] despite normal to very high caloric intake [28-30]. It is correlated with disease progression [31] and patients with higher expanded CAG repeat exhibit more rapid loss of body weight [23]. This phenotype has been evaluated in the different HD mouse models with contradictory results: while R6/2 mice show considerable weight loss, BACHD and YAC128 mice are heavier than wild type mice [32]. In the knock-in Hdh\(^{Q111}\) mice (on a CD1 background), male homozygous showed significantly lower body weight than wild type and heterozygous mice at 28 weeks [32]. To assess the potential effect of the Hdh\(^{Q111}\) mutation (on a B6J background) and of the different genetic background (chr10\(^{B6J/B6J}\) versus chr10\(^{B6J/AJ}\)) we analyzed body weight at three different time stages (3, 14 and 24 weeks). As shown in Figure 5, wild type B6J mice showed significantly lower body weight than wild type B6J.AJ10 mice at 3 weeks of age for both genders (\(p=0.0008\), mean IW±SD: 10.02±1.89 versus 13.03±1.74 for females and \(p<0.0001\), mean IW±SD: 9.22±1.72 versus 13.92±2.29 for males), however, they tended to gain more weight per day than B6J.AJ10 mice in the initial exponential growing phase (\(p=0.0026\), mean EWG±SD: 0.140±0.021 versus 0.107±0.026 for females and \(p=0.0034\), EWG±SD: 0.235±0.039 versus 0.181±0.037 for males). We did not observed any significant difference in weight gain in the final phase between B6J and B6J.AJ10 mice, resulting in similar body weights by week 24 (mean FW±SD: 21.64±1.35 versus 21.99±0.64 for females, and 28.35±1.63 versus 30.13±2.03 in males, respectively). There was no significant difference in weight at any time point between Hdh\(^{Q111}\) mutant and wild type mice, for any of the two different chr10 genetic backgrounds and the same results observed between wild type B6J and B6J.AJ10 mice were found for mice with these backgrounds while carrying a Hdh\(^{Q111}\) mutant allele.

**Nuclear mutant huntingtin.** Another early, dominant, CAG-length dependent phenotype in Hdh\(^{Q111}\) mice is the time-dependent immunostaining of mutant huntingtin in the nuclei of striatal neurons. Previous studies have detected early (~2.5 months) diffuse-immunostaining nuclear mutant huntingtin and later (6–12 months) intranuclear inclusions of mutant huntingtin amino-terminal fragments, using the anti-huntingtin antibody EM48 [13, 14, 22]. In order to determine the effect of the chr10 background on the nuclear mutant huntingtin we immunostained striatal sections from five-months mice with EM48 (striatum and liver; Figure 4D) and the distance between the two modes (liver, Figure 4E).
Results – Article 10

antibody (Figure 6A). As expected, we did not detect any EM48 staining in the wild type mice (Figure 6A, right panel). When comparing Htt knock-in mice we found a similar number of nuclei with diffuse mutant huntingtin (data not shown) while we could detect already some nuclei with intranuclear inclusions of mutant huntingtin amino-terminal fragments (Figure 6A, left panel). We found that, compared to HdhQ111/+B6J mice, the percentage of EM48-positive nuclei with intranuclear inclusions was higher in HdhQ111/+B6J.AJ10 mice (35.88±11.45 versus 46.69±12.11, respectively), although the difference did not reach statistical significance (Figure 6B; p=0.069). Supporting this observation, the number of intranuclear inclusions was found to depend on the constitutive CAG repeat length of HdhQ111/+B6J.AJ10 mice, with longer CAG lengths resulting in an increased percentage of EM48 nuclei with inclusions (and higher when compared with HdhQ111/+B6J mice with similar CAG repeat length) (Figure 6C).

**DARPP-32 and medium spiny neurons loss.** The striatum is one of the key brain areas affected in HD patients, where there is a selective and progressive loss of MSNs [33, 34]. DARPP-32, a dopamine D1 receptor activated molecule that regulates phosphatase and kinase activity, is a marker of these striatal neurons. Levels of DARPP-32 were shown to be substantially reduced in the striatal MSNs of multiple HD mouse models [35-38]. To assess the potential effect of the Htt mutation (on a B6J background) in the levels of DARPP-32 we immunostained striatal sections of 5 months of age mice with DARPP-32 antibody (Figure 7A). We found that DARPP-32 immunoreactivity was significantly reduced (p<0.0001) in the striatum (and nucleus accumbens) of HdhQ111/+B6J (134.2±7.1) by 14% when compared with Hdh+/+B6J control littermates (156.0±3.7) (Figure 7B). The same was observed in mice carrying one copy of the AJ chr10, with HdhQ111/+B6J.AJ10 (137.4±3.7) mice showing a 9% reduction (p=0.0003) when compared with control littermates (150.7±7.0). However, we did not find any significant effect of the chr10 background in the levels of DARPP-32 when comparing B6J with B6J.10 mice, for both wild type (p=0.0920) and mutant (p=0.2521) Htt genotypes. It should be noted though, that Hdh+/+B6J.AJ10 showed a close to significant 3.5% decreased of DARPP-32 levels when compared with wild type B6J mice, and that despite this difference when compared to wild type, HdhQ111/+B6J.AJ10 showed similar levels of DARPP-32 immunoreactivity as HdhQ111/+B6J.

**Discussion** The length of the HD CAG repeat is the single most critical determinant of HD pathogenesis, however the disease expression and its timing are clearly modifiable by other factors. The search for these modifiers has been mostly driven by a candidate gene approach, where genes connected to pathways thought to be involved in HD
pathogenesis are individually tested for an effect on AO. However, the results have been inconsistent and have failed so far to reveal a specific mechanism by which the genetic variation has its apparent effect. Recent advances in genetic technologies are allowing a fast and affordable overview of the human genetic variation in a large set of individuals, setting up the stage for unbiased scans for human genetic modifiers. While genome-wide association studies for HD modifiers are still undergoing, initial scans that relied on genetic linkage to search for chromosome regions associated with AO have already identified a possible region of genetic linkage at the 6q chromosome [8, 9, 39]. However, this genomic region is quite large and has not yet yield specific modifier genes responsible for the effect observed. Instead of examining every single gene within this large region and/or generating a series of transgenic mice to blindly look for potential variants and/or mechanisms that might explain the effect observed at the 6q region, we used a novel approach, a chromosome substitution mouse strain. Previous studies have shown that in HD knock-in mice a variety of phenotypes are modifiable by genetic background [22], and therefore we hypothesized that different chr10 genetic backgrounds (that include homologous genes of those in the human 6q23-24 region) might provide independent evidence to verify the human linkage data.

We generated two sets of HdhQ111 knock-in mice with identical CAG repeat lengths (~141/142±3 CAGs) and genetic background (B6J), with the exception of chr10 in the B6J.AJ10 mice that carried one chromosome from the AJ strain. Both sets of HdhQ111 knock-in mice, independent of their chr10 background, presented a more robust and earlier onset of previously described phenotypes - tissue somatic instability and neuronal intranuclear inclusions (NIIs) - when compared to our original knock-in mice, which had a different background (CD1) and smaller CAG repeats sizes (~111 CAGs). These findings are consistent with the well-established correlation of these phenotypes with CAG repeat length. It cannot be completely excluded that the early manifestation of these phenotypes can be, at least partially, driven by variance of disease penetration on the different background strains. However, this seems unlikely, as our data is consistent with what has been observed for another knock-in mice line with 140 CAGs (on a 129Sv/C57BL6 mix background): no abnormal weight loss and striatal Htt nuclear staining beginning at 8 weeks and nuclear and neuropil aggregates prominent at 16 weeks [40].

In the present study, we observed high levels of repeat expansion in the striatum and liver (instability index of ~8 and ~12, respectively), in concordance with previous studies in other HD knock-in mouse models [12, 41, 42]. We showed that the different genetic background does not seem to have an effect on striatal instability but the AJ chr10 might have a mild detrimental effect on the levels and pattern of repeat instability in the liver.
Even though, the difference did not reach statistical significance, the same trend was observed in all traits analyzed for this tissue (instability index, distance to the longest peak and to the second mode) and was also detectable and in concordance with data from our pilot study with a smaller number of animals (see Supplementary Data). In a recent study, Lee et al. [16] showed that the dynamics of the Htt CAG expansion in the liver and striatum are different, with different mechanisms of expansion and/or cell types probably accounting for this difference. This suggests that different factors are predicted to contribute to liver and striatal instability and that, if real, the effect of the AJ chr10 background is a specific factor involved in the CAG expansion mechanism of the liver but not the striatum. These authors have also shown that these unstable repeats in the liver are highly enriched in polyploidy hepatocytes and that the resultant bimodal distribution of repeat lengths can be observed in HdhQ111/+ mice of different backgrounds indicating that qualitatively similar liver instability can occur across a number of backgrounds [16]. They did not assess instability in knock-in mice with an AJ background, the donor strain used to generate the CSS10 mice used in this study.

One peripheral effect of HD is the loss of body weight despite elevated appetite and caloric intake, with numerous clinical studies observing an impaired energy metabolism in HD patients. A recent study on HD repeat instability has shown that the expression of metabolism genes is actually correlated with liver instability [15]. Based on these findings and our preliminary data on liver instability (see Supplementary data) we analyzed body weight as a way of assessing energy metabolism in our knock-in mice and its possible correlation with the instability observed in the liver. We could not observe any significant difference in body weight between our wild type and heterozygotes mice by 24 weeks of age for either males or females. These findings are consistent with a recent study in an HdhQ111 knock-in mouse model (on a CD1 background), where they could only detect body weight differences by 28 weeks of age and only in male homozygous mice [32]. In contrast to the observations in humans, where weight loss is observed even in HD presymptomatic carriers [43], this study suggests that weight changes may only be detectable in older knock-in heterozygous mice, though we cannot generalize since in our study we used mice with a different genetic background. Indeed, our data showed that there are body weight and therefore metabolism differences between different genetic backgrounds, as we found that B6J mice (independently of the Htt mutation and gender) gained weight more rapidly than B6J.AJ10 mice during the exponential growth phase, and even though the later had a higher initial weight by 24 weeks the two different knock-in mice had similar weights. A phenotype screening of the CSS mice panel [17, 44] has shown that B6J mice tended to be slightly larger at weaning and gained weight much
more rapidly than AJ mice while CSS10 weight gain was intermediate between the parental strains (B6J and AJ mice). It should be noticed though, that these mice were on a high-fat diet while the mice in our study were on a regular diet.

While mutant huntingtin is widely expressed and exerts its toxic effect in many brain regions and peripheral tissues, such as liver, MSNs in the striatum are the most vulnerable in HD [34, 45, 46]. These neurons show a prominent expression of intranuclear (and also cytoplasmic) inclusions containing aggregated expanded repeat huntingtin and other proteins [47-49] and are markedly lost during the progress of the disease. Our data suggests that, even though we could not see any effect of the AJ chr10 on striatum instability, Hdh\(^{Q111/+}\)B6J.AJ10 mice seem to have a slightly higher number of NIs than knock-in mice with a complete B6J background. Even though this difference did not reach statistical significance, it points again for a potential mild effect of the AJ chr10, though not related to the mechanism involved in the striatum CAG repeat instability. The potential outcome of the observed NIs increase is however controversial, while it may represent a protective cellular mechanism of these neuronal cells to attenuate the deleterious effect of the soluble huntingtin protein [50] it can also have a deleterious effect because it is sequestering essential molecules, such as transcription coactivators, and interfering with CBP-regulated gene transcription [51]. On the other hand, we could not find any significant difference in the levels of DARPP-32 between Hdh\(^{Q111/+}\)B6J and Hdh\(^{Q111/+}\)B6J.AJ10 mice, suggesting that levels of neuronal loss in the striatum are similar between these mice, despite the likely increased number of neuronal inclusions on Hdh\(^{Q111/+}\)B6J.AJ10 mice.

Previous studies have shown that DARPP-32 levels are reduced in genetic models of HD compared to wild type littermates, namely R6/2 [35], R6/1 [38], YAC128 [37] and Hdh\(^{Q140/+}\) knock-in mice [36]. We observed a 14% decrease of DARPP-32 immunoreactivity at 5 months of age, when comparing to wild type (on a B6J background), while Hickey et al. showed that DARPP-32 immunoreactivity was normal at 4 months in the striatum of Hdh\(^{Q140/+}\) mice but was reduced by 41% at 1 year [36]. It is noteworthy, that even though we also observed a 9% reduction of DARPP-32 levels on Hdh\(^{Q111/+}\)B6J.AJ10 mice when compared to their wild type littermates, this reduction was not as accentuated as the observed for Hdh\(^{Q111/+}\)B6J. Clearly, the effect of the Htt CAG mutation is higher than the effect of the difference in genetic background (of a single copy of chr10). DARPP-32 is a fundamental component of the dopamine-signaling cascade and the loss of this phosphoprotein may contribute to reported alterations in dopaminergic signaling in transgenic R6 [35, 36, 38, 52] and Hdh\(^{Q150}\) knock-in [53] HD models. Further studies are needed in order to get a more detailed characterization of the DARPP-32 levels, including
transcriptional levels, and its potential effect on dopaminergic signaling pathway, in our HdhQ111 knock-in mouse model.

In summary, the results of our study suggested a potential (but not significant) mild effect of chr10 genetic variants on some of the HD mechanisms, in particular those involved in liver somatic instability and nuclear huntingtin aggregation. In order to confirm these mild effects/trends large cohorts of mice or smaller numbers of very old mice may be required to ensure full penetrance of these phenotypes [14]. Even so, with this approach, QTLs are assigned to the entire chromosome and we cannot distinguish among multiple QTLs on the substituted chr10 [10]. Therefore, these mild effects can be attributed to one or more of the genetic variants reported between B6J and AJ at this chromosome. However, validation and fine-structure mapping of these QTLs is possible by use of other methods, such as sub-strain chr10 consomic mice with region syntenous to 6q23-24 that harbors about 400 non-intronic B6J-AJ variants and/or knock-down or overexpression of particular variants that differ between B6J and AJ for the 71 genes in the 6q syntenous region. By using this chromosome substitution mouse strain approach, we believe that systematic screen for modifiers in HD knock-in mice might help guide the unbiased human genome scan studies that are undergoing.

Acknowledgments EMR is the recipient of a scholarship from the Fundação para a Ciência e a Tecnologia of Portugal (SFRH/BD/44335/2008). Study funding supported by NIH grant from the NINDS NS32765.

Conflict of Interest The authors declare that they have no conflict of interest.

References


In supplementary data:


**Figures**

**Figure 1.** *HdhQ111* knock-in mice with different chr10 genetic backgrounds. Representative diagram of the breeding scheme showing the manner in which both chr10 - AJ or B6J - and *Htt* alleles – wild type (+) or *HdhQ111* knock-in allele (Q^{111}) - were passed to F1 progeny mice used in this study. The actual CAG size of the *HdhQ111* knock-in allele in the paternal mice was 139 CAGs while in the F1 progeny ranged from 133 to 149 CAGs.
Figure 2. Genetic architecture of the ~131 Mbp mouse chr10. (A) Representative homology scheme of mouse chr10 compared with human. (B) Location of the reported genetic variants between B6J and AJ strains at mouse chr10. Coding non-synonymous variants are represented in red, coding synonymous in dark blue, locus-region in orange, mRNA-UTR in purple, splice-site in light blue and intronic in grey.

Figure 3. CAG repeat size of the knock-in alleles in Hdh\textsuperscript{Q111/+}B6J and Hdh\textsuperscript{Q111/+}B6J.AJ10 progeny. The mean CAG repeat size of the mice used in this study was 142.0±3.2 for Hdh\textsuperscript{Q111/+}B6J (n=27) and 141.3±2.8 for Hdh\textsuperscript{Q111/+}B6J.AJ10 (n=28). Solid line represents mean of CAG repeat for each genotype while the dashed line represents the CAG repeat size transmitted from the parental Hdh\textsuperscript{Q111/+}B6J mice (139 CAGs).
Figure 4. Hdh\textsuperscript{Q111/+}\textsuperscript{B6J.AJ10} mice showed a mild increase of liver somatic repeat instability when compared to Hdh\textsuperscript{Q111/+}\textsuperscript{B6J} mice. (A) GeneMapper traces of PCR-amplified Htt CAG repeats from tail, striatum and liver of representative 5 months Hdh\textsuperscript{Q111/+}\textsuperscript{B6J} (144 CAGs) and Hdh\textsuperscript{Q111/+}\textsuperscript{B6J.AJ10} (141 CAGs) mice, and (B) from an oddly unstable Hdh\textsuperscript{Q111/+}\textsuperscript{B6J} (139 CAGs) female mice. (C) Somatic repeat instability was quantified from GeneMapper traces by determining an instability index for tail, striatum and liver of each mouse. Additionally, to capture the different patterns of repeat instability in liver and striatum, we measured (D) the distance between the constitutive repeat mode and the longest repeat after background correction and (E) the distance between the modes of the constitutive and somatically expanded repeats. Hdh\textsuperscript{Q111/+}\textsuperscript{B6J} (n=8) mice are represented in squares and Hdh\textsuperscript{Q111/+}\textsuperscript{B6J.AJ10} (n=8) mice in triangles.
Figure 5. One single copy of AJ chr10 has an effect on the patterns of weight gain. Scatter plots represent the individual values of the initial weight (at ~3 weeks of age), the mean weight gain per day during the exponential growing phase (EWG, up to ~72 days of age) and the mean weight gain per day (WG, up to ~5 months of age) for (A) female and (B) male mice. The mice used for the body weight analysis are as follows: Hdh^{+/+}B6J (13 females and 14 males), Hdh^{111/1+B6J (9 females and 16 males), Hdh^{+/+}B6J.AJ10 (10 females and 10 males) and Hdh^{111/1+B6J.AJ10 (15 females and 5 males). Line represents mean of values for each genotype. **** p<0.0001, *** p<0.001, ** p<0.01 and * p<0.05.
Figure 6. Hdh\textsuperscript{Q111/+}B6J.AJ10 mice showed a mild increase of neuronal intranuclear inclusions when compared to Hdh\textsuperscript{Q111/+}B6J mice. (A) Micrographs of striata stained with EM48 and counterstained with thionine of representative 5 months mice with Hdh\textsuperscript{+/+}B6J, Hdh\textsuperscript{Q111/+}B6J (female, 139 CAGs), Hdh\textsuperscript{Q111/+}B6J.AJ10 and Hdh\textsuperscript{Q111/+}B6J.AJ10 (140 CAGs) genotypes. (B) Quantification of the percentage of EM48-positive cells containing an inclusion (line represents mean of values for each genotype) and (C) its correlation with constitutive Htt CAG repeat size (the best fit linear regression line of inclusion/EM48 nuclei by CAG repeat size for Hdh\textsuperscript{Q111/+}B6J is represent with solid line and for Hdh\textsuperscript{Q111/+}B6J.AJ10 with a dashed line). Hdh\textsuperscript{Q111/+}B6J (n=9) mice are represented in squares and Hdh\textsuperscript{Q111/+}B6J.AJ10 (n=9) mice in triangles. Each individual value represents the mean observed on three consecutive striatal sections for each mouse.
Figure 7. Hdh<sup>Q111/+</sup> knock-in mice showed a dramatic decrease of DARPP-32 immunoreactivity when compared to wild type mice. (A) Micrographs of DARPP-32 positive cells in the striata and nucleus accumbens of representative 5 months mice with Hdh<sup>+/+</sup>B6J, Hdh<sup>111/+</sup>B6J (142 CAGs), Hdh<sup>+/+</sup>B6J.AJ10 and Hdh<sup>111/+</sup>B6J.AJ10 (140 CAGs) genotypes. (B) Quantification of DARPP-32 intensity (arbitrary values ranging from 0 to 255). The mice used for the DARPP32 analysis are as follows: Hdh<sup>+/+</sup>B6J (n=8), Hdh<sup>111/+</sup>B6J (n=9), Hdh<sup>+/+</sup>B6J.AJ10 (n=6) and Hdh<sup>111/+</sup>B6J.AJ10 (n=9). Each individual value represents the mean observed on three consecutive striatal sections for each mouse. Line represents mean of values for each genotype. **** p<0.0001 and *** p<0.001.
Supplementary data

Supplementary Material and Methods. For the pilot study, two crosses were set up: a) one male heterozygous Hdh\textsuperscript{Q111} knock-in mouse was crossed with CSS10 female mice in order to generate wild type Hdh\textsuperscript{+/+}B6J.AJ10 (n=13) and mutant Hdh\textsuperscript{Q111/+}B6J.AJ10 (n=10) mice; and b) another male heterozygous Hdh\textsuperscript{Q111} knock-in mouse was crossed with B6J female mice in order to generate wild type Hdh\textsuperscript{+/+}B6J (n=2) and mutant Hdh\textsuperscript{Q111/+}B6J (n=6) mice. The Htt mutant allele was transmitted from knock-in mice with slightly different original CAG repeat lengths (126 and 128 CAG repeats), which originated a different range of CAG repeat lengths in the Hdh\textsuperscript{Q111/+}B6J (131-137 CAGs, mean 131.8±1.9) and Hdh\textsuperscript{Q111/+}B6J.AJ10 (125-132 CAGs, mean 127.8±3.3) progeny. At 5 months of age, these mice were perfused with periodate-lysine-paraformaldehyde as previously described [13]. One hemisphere of the brain was embedded for sectioning and the other was used for dissection of striatum for analysis of somatic instability. DNA extraction, Hdh\textsuperscript{Q111} knock-in allele genotyping and analysis of somatic instability from selected tissues (tail, striatum and liver) were performed as described in the Material and Methods section of this manuscript. For detection of diffuse nuclear huntingtin, immunostaining with mouse anti-huntingtin monoclonal antibody mAb5374 (Chemicon) was performed on 7μm paraffin-embedded coronal sections of brains hemispheres as previously described [54]. For double staining with EM48/Hoechst, sections were incubated with Hoechst (1:10.000 in TBS) for fifteen minutes before they were mounted with Prolong fade reagent (Invitrogen). Fluorescent microscopy was performed with a Zeiss Axioskop2 microscope equipped with AxioCamMRm camera and AxioVision 4.6 image acquisition software. Four micrographs (20x objective) were taken, with the same exposure time, from the striatum of each mouse. ImageJ image analysis software [19] was used to quantify the number of mAb5374-positive nuclei and the total number of nuclei (as determined by the number of Hoechst-positive nuclei). In order to control for potential differences in cell density we normalized the number of mAb5374-positive nuclei against the number of nuclei.

Supplementary Results Dominant and CAG length-dependent phenotypes, such as Htt somatic instability [12, 15, 16] and nuclear mutant huntingtin [14], were previously described in Hdh\textsuperscript{Q111} knock-in mice with CD1 background. In order to assess the reproducibility of these two phenotypes in Hdh\textsuperscript{Q111} knock-in mice with B6J background we
first examined these phenotypes in a small number of animals. At 5 months of age, we could detect tissue-specific somatic instability, with significant accumulation of expansions in striatum and liver (Supplementary Figure 1), and diffuse immunostaining of mutant huntingtin in the nuclei of striatal neurons (Supplementary Figure 2), in our heterozygotes Hdh\(^{Q111/+}\) mice with B6J background. As shown in Supplementary Figure 1, the levels of somatic instability in the striatum were numerically lower in Hdh\(^{Q111/+}\)B6J (7.63±1.45, n=5) than Hdh\(^{Q111/+}\)B6J.AJ10 (8.53±0.80, n=5), and the same trend was observed in liver, where Hdh\(^{Q111/+}\)B6J (7.23±1.78, n=6) were slightly less unstable than Hdh\(^{Q111/+}\)B6J.AJ10 (8.13±1.07, n=10) mice. On the other hand, we found that the percentage of EM48-positive nuclei was slightly higher in Hdh\(^{Q111/+}\)B6J mice when compared to Hdh\(^{Q111/+}\)B6J.AJ10 mice, respectively, 68.61±4.30 (n=3) versus 61.55±2.39 (n=4), although the difference did not reach statistical significance.

**Supplementary Table 1.** List of mice used for each phenotype assay and their characteristics (group, gender and \(Htt\) CAG repeat).

<table>
<thead>
<tr>
<th>Mouse</th>
<th>Chr10</th>
<th>Gender</th>
<th>(Htt) CAG repeat</th>
<th>Body Weight</th>
<th>Somatic Instability</th>
<th>EM48</th>
<th>DARPP-32</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1558</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td>N.A.</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1853</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td>N.A.</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>R1441</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td>N.A.</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>R1124</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>N.A.</td>
<td>N.A.</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>R1444</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>R1442</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>N.A.</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>926</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>929</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>931</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>932</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1555</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1557</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1559</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1855</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1125</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>N.A.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1126</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>N.A.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1127</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>N.A.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1562</td>
<td>B6J/B6J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1566</td>
<td>B6J/B6J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td>N.A.</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1947</td>
<td>B6J/B6J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td>N.A.</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>R1449</td>
<td>B6J/B6J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td>N.A.</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>R1453</td>
<td>B6J/B6J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td>N.A.</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>R1446</td>
<td>B6J/B6J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Q1563</td>
<td>B6J/B6J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1862</td>
<td>B6J/B6J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1866</td>
<td>B6J/B6J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1948</td>
<td>B6J/B6J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1953</td>
<td>B6J/B6J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>Chr10</td>
<td>Gender</td>
<td>Hit CAG repeat</td>
<td>Body Weight</td>
<td>Somatic Instability</td>
<td>EM48</td>
<td>DARPP-32</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>--------</td>
<td>----------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>------</td>
<td>----------</td>
</tr>
<tr>
<td>Q1956</td>
<td>B6/J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1196</td>
<td>B6/J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1197</td>
<td>B6/J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1455</td>
<td>B6/J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>928</td>
<td>B6/J</td>
<td>F</td>
<td>139</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>934</td>
<td>B6/J</td>
<td>F</td>
<td>139</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>930</td>
<td>B6/J</td>
<td>F</td>
<td>142</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>933</td>
<td>B6/J</td>
<td>F</td>
<td>143</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1854</td>
<td>B6/J</td>
<td>F</td>
<td>145</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1556</td>
<td>B6/J</td>
<td>F</td>
<td>139</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1856</td>
<td>B6/J</td>
<td>F</td>
<td>142</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1945</td>
<td>B6/J</td>
<td>F</td>
<td>143</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>R1123</td>
<td>B6/J</td>
<td>M</td>
<td>133</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>R1451</td>
<td>B6/J</td>
<td>M</td>
<td>136</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1861</td>
<td>B6/J</td>
<td>M</td>
<td>144</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1949</td>
<td>B6/J</td>
<td>M</td>
<td>144</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1988</td>
<td>B6/J</td>
<td>M</td>
<td>145</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1864</td>
<td>B6/J</td>
<td>M</td>
<td>142</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1867</td>
<td>B6/J</td>
<td>M</td>
<td>142</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1565</td>
<td>B6/J</td>
<td>M</td>
<td>143</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1858</td>
<td>B6/J</td>
<td>M</td>
<td>143</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1865</td>
<td>B6/J</td>
<td>M</td>
<td>146</td>
<td>N.A.</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>R1454</td>
<td>B6/J</td>
<td>M</td>
<td>139</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1450</td>
<td>B6/J</td>
<td>M</td>
<td>140</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1447</td>
<td>B6/J</td>
<td>M</td>
<td>143</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1448</td>
<td>B6/J</td>
<td>M</td>
<td>143</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1950</td>
<td>B6/J</td>
<td>M</td>
<td>144</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1452</td>
<td>B6/J</td>
<td>M</td>
<td>146</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1989</td>
<td>B6/J</td>
<td>M</td>
<td>147</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1683</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td>N.A.</td>
<td>Yes</td>
</tr>
<tr>
<td>Q1757</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td>N.A.</td>
<td>Yes</td>
</tr>
<tr>
<td>Q1758</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td>N.A.</td>
<td>Yes</td>
</tr>
<tr>
<td>R1467</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>R1494</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Q1679</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1756</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1759</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1958</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1497</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1794</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td>N.A.</td>
<td></td>
</tr>
<tr>
<td>Q1795</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td>N.A.</td>
<td></td>
</tr>
<tr>
<td>Q1796</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td>N.A.</td>
<td></td>
</tr>
<tr>
<td>Q1797</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td>N.A.</td>
<td></td>
</tr>
<tr>
<td>R1602</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td>N.A.</td>
<td></td>
</tr>
<tr>
<td>R1603</td>
<td>B6/J</td>
<td>F</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td>N.A.</td>
<td></td>
</tr>
<tr>
<td>Q1684</td>
<td>B6/J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td>N.A.</td>
<td>Yes</td>
</tr>
<tr>
<td>R1605</td>
<td>B6/J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td>N.A.</td>
<td>Yes</td>
</tr>
<tr>
<td>R1920</td>
<td>B6/J</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td>N.A.</td>
<td>Yes</td>
</tr>
<tr>
<td>Mouse</td>
<td>Chr10</td>
<td>Gender</td>
<td>Htt CAG repeat</td>
<td>Body Weight</td>
<td>Somatic Instability</td>
<td>EM48</td>
<td>DARPP-32</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>----------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>------</td>
<td>----------</td>
</tr>
<tr>
<td>R1473</td>
<td>B6J/AJ</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>R1488</td>
<td>B6J/AJ</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Q1686</td>
<td>B6J/AJ</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1760</td>
<td>B6J/AJ</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1761</td>
<td>B6J/AJ</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1762</td>
<td>B6J/AJ</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1868</td>
<td>B6J/AJ</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1869</td>
<td>B6J/AJ</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1870</td>
<td>B6J/AJ</td>
<td>M</td>
<td>WT</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1798</td>
<td>B6J/AJ</td>
<td>M</td>
<td>WT</td>
<td>N.A.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1799</td>
<td>B6J/AJ</td>
<td>M</td>
<td>WT</td>
<td>N.A.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1607</td>
<td>B6J/AJ</td>
<td>M</td>
<td>WT</td>
<td>N.A.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1917</td>
<td>B6J/AJ</td>
<td>M</td>
<td>WT</td>
<td>N.A.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1495</td>
<td>B6J/AJ</td>
<td>F</td>
<td>138</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1493</td>
<td>B6J/AJ</td>
<td>F</td>
<td>139</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1959</td>
<td>B6J/AJ</td>
<td>F</td>
<td>140</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1753</td>
<td>B6J/AJ</td>
<td>F</td>
<td>144</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1601</td>
<td>B6J/AJ</td>
<td>F</td>
<td>141</td>
<td>N.A.</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1682</td>
<td>B6J/AJ</td>
<td>F</td>
<td>140</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>R1341</td>
<td>B6J/AJ</td>
<td>F</td>
<td>140</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1754</td>
<td>B6J/AJ</td>
<td>F</td>
<td>141</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1680</td>
<td>B6J/AJ</td>
<td>F</td>
<td>144</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>R1340</td>
<td>B6J/AJ</td>
<td>F</td>
<td>144</td>
<td>Yes</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Q1681</td>
<td>B6J/AJ</td>
<td>F</td>
<td>136</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1865</td>
<td>B6J/AJ</td>
<td>F</td>
<td>139</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1468</td>
<td>B6J/AJ</td>
<td>F</td>
<td>140</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1496</td>
<td>B6J/AJ</td>
<td>F</td>
<td>142</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1863</td>
<td>B6J/AJ</td>
<td>F</td>
<td>143</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1957</td>
<td>B6J/AJ</td>
<td>F</td>
<td>149</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1604</td>
<td>B6J/AJ</td>
<td>F</td>
<td>141</td>
<td>N.A.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1916</td>
<td>B6J/AJ</td>
<td>F</td>
<td>146</td>
<td>N.A.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1866</td>
<td>B6J/AJ</td>
<td>M</td>
<td>137</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1343</td>
<td>B6J/AJ</td>
<td>M</td>
<td>140</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1764</td>
<td>B6J/AJ</td>
<td>M</td>
<td>142</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1344</td>
<td>B6J/AJ</td>
<td>M</td>
<td>143</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1961</td>
<td>B6J/AJ</td>
<td>M</td>
<td>140</td>
<td>N.A.</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1342</td>
<td>B6J/AJ</td>
<td>M</td>
<td>141</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>R1608</td>
<td>B6J/AJ</td>
<td>M</td>
<td>140</td>
<td>N.A.</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>R1918</td>
<td>B6J/AJ</td>
<td>M</td>
<td>140</td>
<td>N.A.</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>R1919</td>
<td>B6J/AJ</td>
<td>M</td>
<td>140</td>
<td>N.A.</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Q1960</td>
<td>B6J/AJ</td>
<td>M</td>
<td>145</td>
<td>N.A.</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: Chr chromosome; N.A. not ascertained; x failed; WT wild type (Hdh+/+).
Supplementary Figure 1. Preliminary data for \( \text{Hdh}^{Q111} \) CAG repeat instability. Somatic repeat instability was quantified from GeneMapper traces by determining an instability index for tail, striatum and liver of each mouse. \( \text{Hdh}^{Q111/+} \) B6J mice are represented in squares and \( \text{Hdh}^{Q111/+} \) B6J.AJ10 in triangles. Line represents mean of values for each genotype.

Supplementary Figure 2. Preliminary data for nuclear mutant huntingtin immunostaining in striatal neurons. Fluorescent micrographs of striata double-stained with mAb5374 and Hoechst of a representative 5 months mouse for each genotype.
Final Considerations
“In other words, there is no obvious clear distinction between simple Mendelian and complex traits: genetic diseases represent a continuum with diminishing influence from a single primary gene influenced by modifier genes, to increasingly shared influence by multiple genes.”


Over 4,000 human diseases are caused by a single mutation in a single gene. These monogenic disorders are a dichotomous trait, in the sense that the presence or absence of the causative mutation determines whether someone will get the disease or not. By contrast, in complex disorders, such as heart disease, diabetes and many psychiatric diseases, is the combination of multiple interacting genes and environmental factors that result in the manifestation of the disease. Even though monogenic disorders are inherited in a "simple" Mendelian way, phenotypes among carriers of the mutation can subtly or profoundly differ in diverse biological features such as AO, severity and other associated phenotypic properties [272]. In many Mendelian disorders, some proportion of phenotype variability can be associated with the nature of the mutation in the disease-causing gene [273], while the remainder variation is due to combined effects of genetic, environmental and stochastic factors that convert “simple” Mendelian disorders into complex traits [274]. The challenge now is to assess the relative contribution of genetic factors distinct from the causing gene that underlie disease variability. As it has been shown in complex disorders, it is expected that these complex traits are more likely to be caused by several genetic variations, each with a small overall contribution making it hard to determine their relative effect.

HD is among this class of monogenic disorders that display wide variability in its clinical features. Though considered as a "simple" Mendelian disorder, observation of alleles with reduce penetrance and the large variation in AO for any given allele size (mean +/- 20 years), even in the fully penetrant range [183], clearly demonstrates that residual AO is a complex trait. HD is a highly disable disorder for which there is still no effective treatment. So, in the paucity of a treatment, HD research is focused in understanding the complex disease mechanism that leads to the overt of symptoms and ultimately unveil effective therapeutic targets to slow the progression or ameliorate HD symptoms.
An attractive hypothesis is that the chromosomal region around the \textit{HTT} gene harbors variation that may lead to the identification of genetic modifiers affecting not only HD pathogenesis but also \textit{HTT} repeat instability. In fact, a detailed examination of frequency and magnitude of \textit{HTT} CAG repeat instability in large normal and reduced penetrance alleles in the general population [Article 1] and full penetrance alleles from HD families [Article 2] of Portuguese origin, suggested the existence of major family-specific instability modifiers. Moreover, early studies have found suggestive evidence for AO [184,275] and CAG instability [180] modifiers near the \textit{HTT} gene, raising a potential role for the 4p16.3 haplotype as a \textit{cis}-acting modifier. Delineation of the haplotype structure of the 4p16.3 region also provides an opportunity for personalized therapeutics in HD through the development of a small number of SNP-targeted antisense oligonucleotide molecules that can selectively silence mutant \textit{HTT} [276] and be sufficient to treat up to 88% of the HD patient population [180].

In our original haplotype analysis of the \textit{HTT} region, which extended previous lower-resolution studies, we defined in HD individuals of Western/Northern European ancestry multiple disease-associated SNP haplotypes, consistent with multiple independent origins of HD in this population [Supplementary Article 1]. Nevertheless, our analysis revealed the presence of a major haplotype (HTT_Hap1) that accounted for ~50% of the disease chromosomes and solely 10% of the non-HD chromosomes. Even though HTT_Hap1 haplotype is likely to be enriched in higher-range normal alleles and therefore act as a reservoir of new mutations to HD, we observed no significant difference in the distribution of expanded CAG repeats. Our detailed haplotype analysis also did not provide evidence of an impact on the AO by the \textit{HTT} haplotypes either in the HD or in the normal chromosomes.

Following up our initial haplotype analysis, we defined the disease haplotype in HD chromosomes from individuals of Southern European ancestry amongst a large set of unrelated multiplex Portuguese HD families [Article 4]. We observed four different disease haplotypes at an elevated frequency: the “major Northern/Western European” haplotype (haplotype A that is closely related to HTT_Hap1 in our initial study) and three other distinct haplotypes. In contrast to the Northern/Western European haplotypes where HTT_Hap1 accounted for ~50% of these HD chromosomes, our closely related haplotype (haplotype A) only accounted for 25% of the Southern European HD chromosomes, where the major haplotype was actually haplotype B, present in about 40% of these chromosomes. These findings overemphasize the established concept that the HD mutation arose multiple times in the European population, giving rise to at least two
distinct major HD haplotypes, haplotype A (“Northern-European”) and haplotype B (“Southern-European”).

In accordance with our original haplotype study [Supplementary Article 1], haplotype A (related to HTT_Hap1) was found to be enriched in high-normal alleles (median normal CAG size was 26.0 compared to 17.0-18.0 in the other haplotypes) [Article 4] and was also shared by the two biggest HD expansions observed in the Portuguese HD families [Article 2]. Despite these findings, we also found evidence of high instability of the expanded repeat in the other “Southern-European” haplotypes, as the biggest reported expanded CAG allele in the Portuguese HD families was observed in haplotype B (the major “Southern-European”). Furthermore, analysis of the HTT CAG repeat in families representing the general population living in the mountain valleys around Bolzano, revealed intergenerational instability of the normal repeat on three non-HD chromosomes with the “Southern European” disease haplotypes. In contrast to recent findings suggesting that there is a major predisposing haplotype that influences CAG instability in HTT [180], we found no clear effect of local variation on the instability of this dynamic repeat. The strength of our study relies on the fact that we assessed intergenerational instability in phased haplotypes directly inferred by family structure.

Our haplotype studies in the Northern/Western and Southern European sub-populations raised the interesting hypothesis that the dynamic HTT CAG repeat might be subject to selection in different populations and that in order to develop therapeutics aimed at the mutant HTT allele it is important to understand the HTT haplotype in the different populations. Furthermore, even though we cannot rule out a possible influence of the 4p16.3 haplotype in other HD phenotypes, our data suggest that genetic modifiers of AO and repeat instability should be sought elsewhere in the genome.

For about two decades now, several genetic factors have been investigated in HD through candidate gene studies, with variants within the HTT gene [187-192] and over other 20 different genes being implicated as modifiers [184,187-195,198-222]. From these, by 2009 when this PhD project started, only a few associations had been reproduced while the majority showed inconsistent and conflicted results across studies.

A major problem of most association studies is the definition of phenotype; with AO being the most commonly used phenotype in HD modifier studies. However, its precise determination is not trivial, especially because it relies on the retrospective assessment of events by the patient and family members, and often cognitive and psychiatric symptoms precede the characteristic motor deficits. Because of that, there are variable definitions of AO being used in HD studies in the literature, with some using the earliest AO, regardless
of the type of symptom that inevitably leads to inconsistent and conflicted results. In order to get a more homogenous and narrowly defined phenotype, AO should be defined as the age at which the first motor signs of HD appeared. To emphasize the importance of a robust phenotype, we have recently shown the profound effect that improper estimation of residual AO (AO corrected for CAG size) phenotype can have on the final result when testing the effects of potential genetic modifiers [Article 3]. We initially observed an apparent significant association, in this particular study for the interaction of normal and mutant HTT alleles, that relied on a single outlier with a mutant allele of 120 CAGs and a normal allele of 11 CAGs (two extreme repeat sizes on both the high mutant and low normal repeat range, respectively) and extremely young AO relative to all others. Just by excluding this single 120 CAG subject from our initial regression model we observed a striking effect on the P value (from \( p = 0.006 \) to \( p = 0.506 \)). This finding, also pointed out to another important limitation of candidate genetic association studies in HD, the lack of a detailed and rigorous statistical analysis. Usually, multiple linear regression models are used to test the effect of genetic modifiers on AO. As with all statistical procedures multiple linear regression analysis rests on fundamental assumptions, including normally distributed error with a mean of zero and constant variance (homoscedasticity). Our analysis showed that only HD cases with CAG alleles within the repeat range of 40 to 53 yield a statistically well-behaved data set with constant variance and normally distributed error [Article 3].

Another major problem of association studies is that significant associations may be found as an artifact of population stratification, where ethnic variation or other confounding factors can lead to significant population differences in marker allele frequencies [277]. Indeed, we have exposed ancestry as a critical factor in HD association studies, where European HD cases poorly matched for genetic background might have lead to false positives in association studies of a common polymorphism in the PPARGC1A gene [Article 5]. We observed differences in the contribution of a given marker (rs7665116) between apparently close related European populations, as Southern European HD cases seemed to be not only genetically (different MAF) but also clinically (motor AO 4-5 years later) different from other European samples. If not properly corrected for genetic ancestry, our model would also have incorrectly pointed this gene as a genetic modifier of HD [Article 5]. Interestingly, the positive associations of PPARGC1A found in the literature [208-210] were primarily contributed by patients from Italy, while in a more recent study using samples from the European Huntington’s Disease Network (EHDN) REGISTRY multinational observational study that closely resembled ours, no association of rs7665116 with AO of first symptom was found [211]. This is consistent with our
interpretation and points out the dramatic effect that unaccounted population stratification can have in association studies type I error rates.

These findings [Article 3 and Article 5] have drawn attention in the field to the problems that can confound HD association studies and to the need for consistent phenotype measures, proper study designs and stringent analyses, as acknowledged in an editorial letter in the Neurogenetics journal [278]. More importantly, they provided the basis for a more robust assessment of the effect of several compelling genetic variants in our collection of HD samples, one of the largest collections available.

Indeed, when we applied these critical study design and statistical standards in our large collection of North-American/European HD samples, we failed to replicate some of our own group original positive association studies. Our lab had previously reported that an interaction between the expanded and unexpanded HTT CAG repeat sizes influenced AO of HD, where in individuals with large repeat sizes (47-83) the increase size of the normal allele seemed to mitigate the expression of the disease [188]. However, in our recent rigorous analysis of over 3,500 HD subjects, we found that neither the normal allele CAG length nor the interaction between expanded and normal alleles influences AO of motor manifestations, which was shown to be determined by a completely dominant action of the longest expanded allele and yet unidentified genetic or environmental factors [Article 3]. These findings are consistent with our detailed haplotype analysis of the HTT region that showed that HTT haplotypes on the normal chromosomes did not have a significant impact on AO [Supplementary Article 1]. In a similar way, we also failed to replicate our initial findings associating the 16 TAA repeat allele in the GRIK2 gene with younger HD AO [200], when tested in additional 2,000 HD samples [Supplementary Article 2]. In this particular case, the apparent positive associations of GRIK2 may also have been due to the small number of samples, as these studies comprised no more than 300 subjects [193,200-203]. We also failed to replicate, in our own collection of European HD samples, the previously reported association of AO with specific polymorphisms in PPARGC1A [208-210], GRIN2A and GRIN2B genes [194,204-206] [Article 5 and Article 6]. Moreover, we tested polymorphisms in the COMT, DRD2, DRD4 and DAT1 genes believed to affect dopaminergic neurotransmission and widely investigated in other neurological disorders, but none of these functional polymorphisms revealed a significant modifying effect on the AO of HD motor symptoms [Article 6].

Even though we could not replicate the significant findings of association of these specific candidate-gene studies, it does not necessarily imply lack of causality. The candidate gene or DNA variant can be associated with different relative risks in distinct
Final Considerations

populations and the non-replication might result from real biological differences [279]. Furthermore, it cannot immediately distinguish if the variant is the cause of the observed effect or if it is in linkage disequilibrium with another variant in the vicinity that is the true functional determinant. When this is the case, different studies might have disparate findings for the same gene if this linkage does not exist or varies across populations [279,280]. Interestingly, a meta-analysis including 379 genetic association studies of 36 diseases or traits, found that association studies of the same disease are often inconsistent and that the first study often indicates a stronger effect than what it is seen in the subsequent studies [281].

There is currently large evidence in the literature of CAG repeats known to cause spinocerebellar ataxia (SCA) disorders contributing to complex traits and even acting as modifiers of other diseases. CAG repeat length in the TATA box-binding protein (TBP) gene, known to cause SCA17 when expanded, was found to be associated with risk, AO and prefrontal function in schizophrenia [282,283], while the PPP2R2B gene CAG repeat polymorphism known to cause SCA12 was associated with AD susceptibility in the Japanese and Han Chinese populations [284,285]. Moreover, expansions of the CAG repeat in the ATXN2 gene, which cause SCA2, have been associated with increased risk of sporadic ALS [286-292] and progressive supranuclear palsy, suggesting that ATXN2 repeats may predispose to certain types of neurodegeneration [293]. Interestingly, it has been shown that some of the residual AO in SCA2 can actually be explained by polyQ repeats causing other types of SCA. Long normal CAG repeats in the CACNA1A gene, known to cause SCA6, were associated with disease onset earlier than expected, and explained about 6% of the residual variation in SCA2 AO [294]. Based on these findings and in the phenotypic and molecular overlap that also exists between HD and the other polyQ disorders, we hypothesized that long alleles in these polyQ disease genes could also explain some of the residual AO in HD. However, in our cohort of about 1,500 individuals with HD, we did not find an evident genetic association between HD and expanded CAG repeats in any of the six polyQ disease genes tested [Article 7]. It should be noted though, that we did find three HD cases that carried expanded SCA17 alleles and seven HD cases with intermediate ATXN2 expansions (32-33 repeats) of which some presented clinical features that overlap with the SCA2 phenotype. Hence, we cannot rule out a possible genetic interaction of the CAG repeat at the ATXN2 gene with HD, and further studies in larger cohorts are needed as well as testing the potential effect of these SCA2 intermediate alleles in AO of HD.

In contrast, we tested the potential contribution of the functional HTT CAG repeat to several psychiatric and neurodevelopmental disorders with which HD shares clinical
features. The most common psychiatric symptom in HD is depression, with about 40% of HD patients manifesting signs of major depression [295]. Interestingly, our lab has shown that reduced penetrance CAG alleles are over-represented in diagnosed cases of MDD [296]. These findings lead to the hypothesis that dominant effects of the functional HTT repeat might act as a sensitizer to the effects of the risk factors in the general population that contribute to the manifestation of clinical depression. In addition to depressive signs, mania has also been observed in some HD patients, who alternate between continued periods of depression and mania and are therefore often misdiagnosed with BP. Based on these findings, the obvious next question was to determine if the HTT repeat was also sensitizing individuals to the manifestation of BP signs. However, we found no evidence of a contribution of the HTT repeat to DSM-IV diagnosed cases of BP, which had a frequency of HTT CAG alleles associated with HD even lower than what we observed in our control group, implying that this functional repeat is not a shared factor with the etiology of BP, as it seems to be contributing to MDD [Article 9]. On the other hand, ALS, a motor neuron disease, is speculated to share with HD similar neurodegenerative mechanisms along with a few rare reports of HD patients with concurrent motor neuron signs [297,298] as well as ALS patients sporadically manifesting signs of chorea [299,300]. Additionally, proteins mutated in familial ALS cases have been found in HD inclusions (FUS and TDP-43 protein) [301-303] as well as in cytoplasmatic granules of HD patients’ neurons (OPTN protein) [304]. Despite all of these data supporting a potential role of huntingtin in ALS pathogenesis, our results showed that the HTT CAG repeat has no effect on susceptibility to ALS or its phenotypic manifestations (AO, site of onset and survival after onset), implying that the HD repeat is unlikely to contribute to this disorder [Article 8]. Recent data has corroborated our findings, as no significant association between polyQ length and ALS was found in any of the seven polyQ genes tested beyond ATXN2 [305].

Our studies emphasize the need in the HD field for a better estimate of the expected HD disease prevalence in the general population and set up the stage for further surveys of the prevalence of HTT CAG repeat in other psychiatric and neurodevelopmental disorders. Therefore, we are currently working on collecting a control comparison group that truly represents the general population, and since HTT is an essential regulator of neuronal cell development and synaptic function, we are now assessing the CAG allele distribution in a large set of autism spectrum disorder trios.

The inconsistent outcomes of the various genetic association studies reported in HD evince the disparate study designs employed by HD groups worldwide, using different genotyping platforms, number of SNPs, sample size, ethnic groups and phenotypes. In
order to overcome these limiting factors, we are now examining ~2 million SNPs in about 6,000 HD subjects using whole genome genotyping arrays. This genome-wide approach allowed us to cover the entire genome in the largest collection of worldwide HD samples so far, which is an important requirement for the detection of variants that are expected to have small effects on disease risk. We have gathered all the genotype data and are now able to investigate the effects of several variants at the same time, as well as potential interactions between them, that will hopefully give us information on genes involved in the HD pathogenesis that are part of previously unsuspected pathways. This is a very promising approach, as in about 1,600 human GWA studies that have examined over 800 diseases and traits approximately 10,000 SNP-trait associations have been found [306]. From all these GWA studies, just six have examined AO as a trait in several neurological diseases, namely, ALS, depression, AD, BP, Parkinson’s disease (PD) and multiple sclerosis [307-312]. Only the GWA for AO of depression failed to find any variants reaching genome-wide significance (SNP-trait \(p\)-value <1.0 \(\times\) 10\(^{-5}\)) [312], while 2 to 13 SNPs were found to be associated with AO on the other studied neurological disorders (with risk allele frequency in controls varying from 0.06 to 0.048). For these complex traits, the variants found by GWA had low associated risk that accounted for a small part of the missing heritability.

Both the candidate-gene and GWA approaches, assume that the genetic risk is due to predisposing variants relatively common in the population (frequency > 5%). Therefore our studies do not preclude that other DNA variants in the several genes studied or elsewhere in the genome may act as genetic modifiers of AO. In fact, the answer may reside in rare variants and structural variants, such as copy-number variation, inversions and translocations, which are not well captured by GWAs, as well as interactions between genes or between genes and environmental factors [313]. Indeed, for complex traits like autism and schizophrenia several rare and structural variants have been implicated as affecting the risk of these disorders [reviewed in 314,315] and a recent genome-wide gene-environment study identified GRIN2A as a PD modifier gene via interaction with coffee [316].

In addition to whole-genome screenings of common variants associated with AO of neurological HD symptoms, that are currently undergoing, it is also necessary to identify and assess these rarer but potentially higher risk variants. Whole-exome and whole-genome sequencing have rapidly become the approaches used to study rare variation in different neurological diseases [317], but since these variants are, by definition, rare it does require an even larger number of individuals, which has still high costs associated. One strategy to increase efficiency, and at the same time reduce costs, is to sequence...
individuals that are at both ends of a phenotype distribution. Recently, a genetic modifier study for cystic fibrosis, the first to use exome sequencing and an extreme phenotype study design, discovered that rare variants in DCTN4 are significantly associated with earlier and more chronic infections by Pseudomonas aeruginosa [318]. In a similar way, to assess these rare coding variants, which often have more marked functional consequences, in our large cohort of individuals with HD, we are currently sequencing the whole exome of a selected group of discordant HD samples that have extreme phenotypes. These have either very late or very early onset of neurological symptoms, and are defined as samples at the extremes of the AO distribution after accounting for the expanded CAG repeat size. It is expected that the frequency of alleles contributing to the variance in AO are enriched either in one or both the AO extreme groups [318], and so the originally modest sample size can potentially be used to identify novel candidate genes for follow-up genotyping studies in our large cohort of HD samples.

The application of these unbiased screens holds great promise for the identification of novel genetic variants responsible for the variance in HD AO. But, even though gene targeting permits testing of individual DNA variants for functional effects, the efficient unbiased evaluation of genomic regions that may harbor potential modifier loci is challenging. This has spurred the development of the CSS mice, a novel human-mouse cross-species tool for dissecting complex traits [319]. With this approach, chromosomal regions spanning a QTL previously identified through unbiased screens are bred from a donor inbred strain (A/J) onto a recipient inbred strain (C57BL/6J) and any phenotypic difference in the congenic line relative to the recipient strain is assumed to be caused by donor strain alleles within the congenic region. The power of this strategy to study genetic modifiers has been successfully demonstrated in neurofibromatosis type 1 where chr19A/J was shown to affect both timing and penetrance of the growth of different tumor types associated with this completely penetrant but variably expressed disorder [320]. In essence, the action of genetic modifiers is evident when the genetic background affects the expression of a trait [272]. And this has actually been shown in HD knock-in mouse models where three different inbred strain backgrounds (C57BL/6, FVB/N and 129Sv) manifested variable phenotype severities [268]. This study provided evidence for genetic modifiers of HD CAG repeat instability (increased repeat instability in both C57BL/6 and FVB/N backgrounds compared with 129Sv) and striatal-specific phenotypes (fastest accumulation of nuclear mutant huntingtin and formation of NIs in the C57BL/6 background and slowest in the 129Sv), and were a proof of concept for the use of CSS mice to dissect the puzzling results coming out from HD AO unbiased screens.
Final Considerations

Our lab and others have previously used genetic linkage to identify a human chromosomal region associated with AO on the large 6q23-24 interval [243-245]. This region is in syntenous relationship with mouse chr10 that is highly polymorphic between C57BL/6J and A/J strains. The use of the novel CSS approach, and in this specific case of CSS10 mice, allowed us to compare the effect of this particular chromosomal region in these two different inbred strains. Actually, compelling phenotype differences have already been reported between C57BL/6J (the recipient strain) and CSS10 mice, such as behavioral (anxiety, balance and coordination, fear conditioning, involuntary movements and wildness), hearing and nervous system (autonomic, neuromuscular and sensorimotor functions and sensory gating) phenotypes [321-328]. These findings are strong evidence that C57BL/6-A/J variants within the chr10 region are able to affect the expression of different traits. Moreover, in our own dataset, we showed that the presence versus absence of a single chr10A/J affects the rate of body weight gain [Article 10]. This is a very important observation, that in a certain way validates our HD consomic model, as a phenotype screening of the CSS mice panel has shown the presence of a genetic resistance to diet-induced obesity QTL at chr10, where C57BL/6J mice gained weight more rapidly than A/J mice while CSS10 weight gain was intermediate between the parental strains [329,330]. Our results showed that this trait is affected by the genetic background in a dominant manner, as a single copy of chr10A/J is sufficient to alter the rate of body weight gain.

More importantly, we tested whether dominantly acting chr10 A/J-C57BL6/J variants could influence CAG repeat instability and/or the rate of the early disease process that leads to pathology in heterozygous C57BL6/J HdhQ111 knock-in mice. Our results revealed that a single copy of chr10A/J was sufficient to mildly increase the levels of repeat instability in the liver, the number of neuronal huntingtin nuclear inclusions, and also alter the DARPP-32 levels in MSNs [Article 10]. Even though none of these differences reached statistical significant when comparing HdhQ111/B6 and HdhQ111/B6.AJ10 mice, variants influencing onset of HD symptoms are expected to have relatively small effects in the continuous disease process that leads to pathology in our knock-in model, and therefore in order to ensure full penetrance of these phenotypes large and/or older cohorts of mice might be necessary. Indeed, our study seems to be underpowered as, based on our current data, we would need as twice as many mice per group to detect a significant effect of the chr10A/J variants in the traits analyzed. However, the fact that the same trends were observed in our pilot study with an even smaller number of animals [Article 10 – Supplementary data], suggests that there is at least one A/J variant on chr10A/J modulating some of the mechanisms involved in HD pathogenesis.
One of the major limitations of the CSS approach is that QTLs are assigned to the entire substituted chromosome, and therefore we cannot distinguish if the effects observed in our dataset are due to one or multiple A/J variants within chr10. Even if we take in account only non-synonymous coding SNPs, which are more likely to be functional, there are 106 potentially modifier variants at chr10\textsuperscript{A/J}. From these only 19 are located within homologue genes of the human 6q23-24 linkage interval, such as the \textit{Uttn} (utrophin) gene known to participate in post-synaptic membrane maintenance and acetylcholine receptor clustering, and \textit{Sash1} (SAM and SH3 domain containing 1) that interacts with the actin cytoskeleton and stimulates cell-matrix adhesion [331,332]. Even though there are still a substantial number of genes that are polymorphic at chr10 between C57BL/6J and A/J and therefore potential candidate modifiers of the disease process triggered by the \textit{Htt} CAG repeat allele, our suggestive findings encourage a larger study for fine mapping, which can be easily assessed with the now available sub-consomic panel of CSS10 mice. Our results also nurture the use of CSS and sub-strains to address the need for novel strategies to elucidate genetic modifiers of human disease mutations or risk factors, identified in unbiased genetic studies. This novel approach enables to gather additional information about how modifiers in a specific \textit{locus} functionally modulate a particular complex trait subsequently allowing to better test which of the candidate genes in that region is indeed modifying the trait.

In conclusion, the studies performed in this thesis focused on the application of different classical genetic approaches to study genetic factors and their impact on the disease pathogenesis of highly penetrant disorders, using HD as a disease model.

Our current understanding of the pathological process that leads up to the manifestation of HD clinical symptoms is still scarce and remains a major challenge for the HD research community, who’s ultimate goal is the development of effective therapeutic strategies for this devastating disorder. Our critical studies of candidate gene modifiers might have failed to single out common genetic variants whose functional action produces the heritable variation observed in HD, but they have clearly demonstrated the critical repercussions that ancestry and improper estimation of residual phenotype can have on candidate gene association analysis. More importantly, our work has now prompted other HD researchers to take into account these potential sources of bias in their genetic association studies. We believe we have provided the basis for a more rigorous and reliable statistical assessment of candidate modifier variants in HD research, and genetic modifiers research in general, hopefully avoiding a redundant misuse of time, energy, and funds pursuing false leads as targets for pharmacological intervention. The dearth of believable genetic modifiers for AO of HD motor symptoms after 20 years of research...
Final Considerations

highlights the importance of unbiased approaches in HD. We anticipate that the current efforts of the worldwide multi-center GWAS consortium, which we are a major contributor, and of genome-wide screens for common and rare variants will prove to be the most rapid and definitive way to identify genetic modifiers of HD pathogenesis. Ultimately, for therapeutic development, it will be necessary to define the mechanism of action of the candidate regions of the genome implicated by these unbiased screens, which typically are large precluding unbiased gene-by-gene analyses. Our encouragingly data using the novel CSS mice approach, to evaluate the large chromosomal region that consistently emerged from linkage studies in HD families, sets up the stage for the use of this pioneering cross-species functional approach to map and evaluate modifiers by parsing the candidate regions.

In the long run, we envisage that understanding the mechanisms modulating HD disease expressivity will lead to improve prediction, management, treatment and perhaps even prevention of HD.
References

37. Gerfen CR (1992) The neostriatal mosaic: multiple levels of compartmental organization. Trends Neurosci 15 (4):133-139.


42. Gusella JF (2001) Huntington Disease. eLS.


194

Genetic modifiers of Huntington’s disease


References


304. Hortobagyi T, Troakes C, Nishimura AL et al. (2011) Optineurin inclusions occur in a minority of TDP-43 positive ALS and FTLD-TDP cases and are rarely observed in other neurodegenerative disorders. Acta Neuropathol 121 (4):519-527.


Annex
Supplementary Article 1: Common SNP-Based Haplotype Analysis of the 4p16.3 Huntington Disease Gene Region.


Reprinted with the permission of Elsevier
License number 3192510505355
Common SNP-Based Haplotype Analysis of the 4p16.3 Huntington Disease Gene Region


Age at the onset of motor symptoms in Huntington disease (HD) is determined largely by the length of a CAG repeat expansion in HD but is also influenced by other genetic factors. We tested whether common genetic variation near the mutation site is associated with differences in the distribution of expanded CAG alleles or age at the onset of motor symptoms. To define disease-associated single-nucleotide polymorphisms (SNPs), we compared 4p16.3 SNPs in HD subjects with population controls in a case-control strategy, which revealed that the strongest signals occurred at a great distance from the HD mutation as a result of “synthetic association” with SNP alleles that are of low frequency in population controls. Detailed analysis delineated a prominent ancestral haplotype that accounted for ~50% of HD chromosomes and extended to at least 938 kb on about half of these. Together, the seven most abundant haplotypes accounted for ~83% of HD chromosomes. Neither the extended shared haplotype nor the individual local HD haplotype were associated with altered CAG-repeat length distribution or residual age at the onset of motor symptoms, arguing against modification of these disease features by common cis-regulatory elements. Similarly, the 11 most frequent control haplotypes showed no modifier effect on age at the onset of motor symptoms. Our results argue against common local regulatory variation as a factor influencing HD pathogenesis, suggesting that genetic modifiers be sought elsewhere in the genome. They also indicate that genome-wide association analysis with a small number of cases can be effective for local regionalization of genetic defects, even when a founder effect accounts for only a fraction of the disorder.

Introduction

Huntington disease (HD; MIM 143100) is a dominantly inherited neurodegenerative disorder characterized by involuntary movements, motor deficits, cognitive decline, and psychiatric disturbance. The genetic defect was originally mapped to chromosome 4p16.3 by linkage to DNA polymorphisms and then confined to a segment of ~2 Mb by crossover analysis in families. Further narrowing of the candidate region was accomplished by association analysis with both multi-allele and dimorphic DNA markers, which revealed the existence of multiple ancestral haplotypes and led to the identification of the genetic defect as an expansion mutation in the CAG trinucleotide repeat of HDI (formerly HD [MIM 613004]) in chromosomal region 4p16.3. Not only does the length of the HDI CAG repeat determine the probability that HD will be manifest (alleles of 36–39 repeat units do not show full penetrance), but it is also the primary determinant of when these symptoms develop; when there are more than 39 repeats, the age at onset of diagnostic motor symptoms decreases with increasing CAG length. Importantly, some portion of the variance in the age at onset is not explained by the CAG-repeat length. For example, in...
a sample of 492 HD individuals with a CAG repeat length of 44, ~46% developed diagnostic motor symptoms at an age that was more than 5 years earlier or later than the average age at onset (44.4 years). The unexplained variance in age at onset shows heritability, supporting the idea that genetic factors in addition to the CAG repeat are involved in determining the age at onset of HD symptoms. Though genetic modifiers of HD might be located anywhere in the genome, an attractive hypothesis is that the chromosomal region of HTT itself harbors variation that alters HD pathogenesis. Interestingly, even before the expanded CAG repeat was identified, a locus closely linked to the HD mutation was hypothesized to modify age at onset. Subsequently, suggestive evidence was reported for linkage of an age-at-onset modifier near HTT, raising a potential role for the 4p16.3 haplotype surrounding the CAG expansion as a cis-acting modifier of the rate of disease pathogenesis leading to the onset of motor symptoms, possibly through regulation of HTT expression or through an effect on its protein product, huntingtin. Similarly, a cis-acting factor affecting the stability of the CAG repeat has also been proposed as a factor in the disease. Because HD proceeds inexorably to death an average of 16 years after the onset of motor symptoms and because there is no preventative or attenuating treatment, a mechanism capable of delaying the onset of HD symptoms could be of great benefit. Consequently, we have tested the hypothesis of a cis-acting modifier of CAG-repeats length and/or the onset of HD motor symptoms by evaluating the extended SNP haplotypes of the HTT region from mutant and normal chromosomes. In carrying out this study, we also uncovered "synthetic association" signals, which we examined in more detail in order to guide future association analyses of HD modifiers and other disorders.

Material and Methods

Subjects, SNP Genotyping, and Quality Control

DNA samples from HD subjects were collected locally or from collaborating investigators by the Huntington's Disease Center Without Walls at Massachusetts General Hospital with proper informed consent and were studied with approval of the Partners Healthcare Institutional Review Board. SNP genotyping was performed with the Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara CA) at the Broad Institute of MIT and Harvard (Cambridge MA). As population controls, we used individuals who were of European ancestry and who had been genotyped with the same platform in a previous study but who were not assessed for HD. For a stringent analysis, initial quality controls (QC) were applied to each cohort independently (SNP genotyping call rate > 99%; minor-allele frequency > 5%). The Hardy-Weinberg equilibrium p value filter was applied to controls (p > 1 × 10^-6). On the basis of high-quality SNPs that had passed QC, we identified unrelated subjects of European ancestry by comparing estimated identity-by-descent (IBD) calculated by the PLINK program to those of HapMap samples. Subsequently, SNPs showing significantly different call rates between HD subjects and controls were also excluded (p < 0.001). Finally, we obtained genotypes for 436,185 SNPs for 2,375 individuals (699 HD subjects and 1,676 controls) (sample genotyping call rate: 99.903%).

Genotyping of the Delta2642 In/Del Polymorphism

We genotyped delta2642 in HD subjects by using a previously established PCR assay with slight modifications. In brief, genomic DNA (30 ng) was amplified by PCR with fluoroently labeled primers (6-FAM-5’-GCTGGGGAACAGACTCACACC-3’ and 5’-CCTGGAGTTCCGAGACCTTG-3’), then separated with an ABI 3730 DNA analyzer (Applied Biosystems, Foster City, CA). Subsequently, the sizes of resolved PCR products were determined by the GeneMapper program (version 3.7) on the basis of the internal size standards (GeneScan 500-LIZ) so that sample genotypes could be obtained.

Genome-wide Association Analysis

We performed a case-control association analysis that compared genotypes of HD subjects (699 unrelated HD individuals) to those of independent population controls (1,676 individuals not assessed for HD) by using a chi-square test implemented in PLINK. Because HD is dominantly inherited, a dominant model was used, and SNPs with a genotype count smaller than 5 in either HD subjects or controls were excluded. We observed significant association (p < 5 × 10^-8) for multiple markers in chromosomal region 4p16.3 and for six other SNPs (rs3754791, rs1376313, rs4606566, rs1405178, rs1360809, and rs16928445) not in this region. We were able to dismiss the latter as being due to technical artifacts by performing proxy-drop association analysis with the PLINK program. In this analysis, we excluded and then reimputed each SNP, and we used neighboring SNPs for the subsequent association analysis (data not shown). We then extracted all SNPs in the 4p16.3 region of association for further analysis.

Haplotype Phasing and Analysis

Haplotype phasing of SNPs was performed with the MACH program. 4p16.3 SNPs that showed significant association with HD (p < 0.01) in the original dominant-model analysis (59 SNPs) or in the dropout association analysis (21 SNPs) were used for haplotype phasing. Among these, 21 SNPs were significant in both analyses (Table S1, available online). In addition to genotypes of 61 SNPs, we included genotypes of HD CAG-repeat mutation (expanded or normal) and the delta2642 polymorphism (insertion or deletion) in phasing in order to identify HD chromosomes and the major HD haplotype chromosome, respectively, in HD subjects. The genotype of the HTT CAG repeat in HD individuals was coded as a biallelic system, in that each individual was a heterozygote with one expanded and one normal allele. Among controls, each individual was coded as a homozygote with two normal alleles of the HTT CAG repeat. Association of haplotypes with differences in CAG length, age at onset, or residual age at onset of motor symptoms was tested in an ANOVA model followed by post-hoc test with Tukey's HSD test in R (2.7.2). Residual age at onset of motor symptoms of an HD subject was the difference between the actual natural log-transformed age at onset of that individual and the natural log-transformed age at onset predicted on the basis of a large collection of HD subjects with the same expanded HD CAG-repeat length.
Figure 1. Huntington Disease Case-Control SNP Association Analysis

(A) A regional association plot focusing on the 4p16.3 region of significant association signals from HD case-control analysis is shown. Green lines represent genes, and the blue line shows the recombination rate (secondary y axis) based on HapMap release 22 data. Genomic locations of SNPs and genes were based on the UCSC hg18 genome assembly. SNP rs12641989 (red diamond) maps closest to the HD CAG expansion mutation (inverted red triangle). SNP rs12641989 (red diamond) shows the strongest association signal ($p = 1.73 \times 10^{-39}$), and the distant SNP rs11248108 (green diamond) also showed strong association with the disease ($p = 1.74 \times 10^{-22}$). (B) To test whether rs12641989 tagged the most abundant HD haplotype and to identify SNPs associated with other haplotypes, we compared the association signals in the original association analysis (699 HD versus 1676 controls; red trace) to those from an association analysis (315 HD versus 1676 controls; blue trace) from which HD samples with at least one minor allele for rs12641989 had been excluded. The location of the HTT CAG repeat is given as an inverted red triangle. Above the map are bars indicating two segments for which detailed haplotype analysis was performed: EXT_Hap (comprising 62 markers from rs11248108 to rs12641989) and HTT_Hap (comprising 21 markers spanning HTT) (for markers see Table S1). SNPs from the former were arbitrarily numbered sequentially from telomeric to centromeric (e.g., 1, rs1248108; 10, rs189139; 12, rs762847; 24, rs2857845; 28, rs2285086; 34, rs11731237; 36, rs2259869; 39, rs82333; 41, delta2642 (black triangle); 50, rs1136890; 56, rs1730768; and 62, rs12641989).

Results

SNP Comparison of HD and Control Subjects

HTT maps approximately 3 Mb from the short-arm telomere in the p16.3 subband of chromosome 4. As a first step to examining the haplotype structure of HD chromosomes, we extracted genotypes from this region for 699 unrelated HD subjects of European origin and 1,676 population controls. We generated these genotypes by using the Affymetrix Genome-Wide Human SNP Array 6.0 platform and compared them by standard case-control association analysis (see Figure S1 for a Q-Q plot). Within a -1 Mb region, 40 SNPs showed genome-wide significance (dominant model, $p < 5 \times 10^{-8}$) for association with the disorder, and most of these were centered as expected around HTT (Figure 1A). Surprisingly, however, the SNP most significantly associated with HD (i.e., rs12641989; red diamond; $p = 1.73 \times 10^{-39}$) did not map to the CAG expansion mutation. To confirm this interpretation, we excluded HD subjects with at least one minor allele for rs12641989 (384 subjects) and performed the case-control association analysis again by using the remaining HD subjects (315 subjects) and all controls (1,676 subjects). The plot in Figure 1B shows the results of the association analyses (dominant model) before (red trace) and after (blue trace) the exclusion of
HD subjects with the minor allele of rs12641989 (marker 62). As expected, many SNP association signals were attributable to HD-mutation-bearing chromosomes marked by the rs12641989 minor allele; these signals disappeared when these HD chromosomes were excluded. Surprisingly, among the association signals that disappeared was the one at rs11248180 (marker 1), indicating that both rs12641989 and rs11248180 mark an extended haplotype of ~938 kb associated with the most abundant HD-mutation-bearing chromosome. Notably, the exclusion of HD samples on the basis of rs12641989 also generated new genome-wide-significant association signals at rs2857845 (marker 24) and rs1730768 (marker 56), indicating that these markers tag one or more haplotypes whose ancestral origin is likely to be different from that of the most frequent HD haplotype. These findings are consistent with the multi-allele-marker haplotype studies that led originally to localization of the HD mutation and which revealed that although the HTT region exhibits many different haplotypes, one apparent founder haplotype, marked by an HTT-codon deletion polymorphism, delta2642, was particularly frequent among HD-mutation-bearing chromosomes (for convenience, here we refer to this haplotype as the “major” HD haplotype).16,19

SNP-Based Haplotypes across the HTT Region
To determine the proportion of HD chromosomes bearing the extended 938 kb haplotype detected above, we generated SNP haplotypes (including the delta2642 codon deletion) for 62 polymorphic sites from rs11248180 to rs12641989 (Figure 1B; Table S1). We assumed that each individual with HD had one mutant (defined as having >35 CAGs) and one wild-type (defined as having <36 CAGs) allele. Fully 165 (23.6%) of the 699 HD chromosomes shared the entire 62 SNP haplotype (Table S2), whereas only 2.1% of control chromosomes did so, suggesting a common ancestral origin for these HD chromosomes and preservation of an extended region of identity by descent due to the relatively low recombination rate across this segment of 4p16.3 (Figure 1A, light blue trace).

To test for evidence of a cis-acting modifier within this large segment, we compared HD subjects who shared this haplotype to those with all other HD haplotypes across this segment. There was no significant difference in the distribution of expanded CAG-repeat lengths, the distribution of ages at onset of diagnostic neurologic manifestations in these subjects, or the distribution of residual age at onset after the effect of the CAG-repeat length was taken into account (Figure 2A).

Although the extended 938 kb haplotype was readily detectable, the even greater frequency of the delta2642 deletion allele suggested that this same ancestral chromosome was also the source of additional HD chromosomes among the 699 whose shared haplotypes, as a result of historical recombination events, were not as extended. Consequently, we generated SNP haplotypes spanning HTT for both HD and normal chromosomes and compared all HD-mutation-bearing haplotypes that occurred 20 or more times among the 699 HD chromosomes. This more restricted haplotype of 21 markers spanned ~235 kb from rs2857845, 48 kb upstream of HTT, to rs3095073, 17 kb downstream (Figure 1B; Table S3). The first three SNPs were located telomeric to the site of the CAG repeat, and the remainder were centromeric to it.

As expected, the vast majority of chromosomes bearing a CAG expansion could be grouped into a limited number of HD haplotypes; ~83% of all HD chromosomes fell into one of seven different haplotypes (Figure 3). Of these, the most frequent, HTT_Hap_1 (which forms the core of the shared extended haplotype noted above), accounted for almost half of all HD-mutation-bearing chromosomes. Most of the remaining haplotypes are likely to have arisen from the occurrence of independent HD mutations, although some could have been derived from HTT_Hap_1 or from one another by one or more historical recombination events within the 235 kb segment spanned by this haplotype. For example, HTT_Hap_5, present on ~5% of HD-mutation-bearing chromosomes, might have been derived from HTT_Hap_1 as a result of a rare recombination event between rs2857845 and rs2471347, the ultimate and penultimate SNPs on the telomeric end of the haplotype (Figure 3B).

For each of these seven most frequent HD haplotypes, we again examined the distribution of CAG repeat lengths on HD-mutation-bearing chromosomes, the distribution of ages at onset of these subjects, and the distribution of residual age at onset after we had accounted for the effect of the CAG-repeat length. By ANOVA, we found that the individual HD haplotypes did not significantly influence CAG-repeat length, age at onset, or residual age at onset (Figure 2B). We also compared the subjects with HTT_Hap_1 against all other HD subjects and again found no significant difference in any of the three parameters (data not shown).

Thus, neither HTT nor the almost 1 Mb surrounding region represented by the common extended HD haplotype showed evidence of a cis-acting modifier of either CAG-repeat length or age at onset of motor symptoms.

Lack of HD Modification by the Normal-Chromosome HTT Region
The haplotype structure of the normal chromosome in each HD individual also offered us the potential to test whether the haplotype of the normal HTT allele acts as a modifier of the age at onset of motor symptoms. We have previously showed that the length of the HTT CAG repeat on the normal chromosome, and by consequence the length of the polyglutamine tract in normal huntingtin, does not modify the age at onset of motor symptoms caused by mutant huntingtin.7 However, although normal huntingtin’s polyglutamine tract does not have an impact, it is conceivable that differences that occur in the level or pattern of expression of the normal HTT allele and which
Figure 2. Local 4p16.3 Haplotype Does Not Influence Distribution of Either HD Expanded CAG Repeat Length or the Age at the Onset of Motor Symptoms

(A) The distributions of expanded CAG repeat length (top), age at the onset of motor symptoms (middle), and residual of age at the onset of motor symptoms after the length of the CAG repeat (bottom) had been taken into account were compared for 165 individuals with an identical extended haplotype on their HD-mutation-bearing chromosome (1: Extended haplotype 1) across the EXT_Hap segment (Figure 1B) versus the other 424 HD subjects (Extended haplotype Other).

(B) The distributions of expanded CAG repeat length (top), age at the onset of motor symptoms (middle), and residual of age at the onset of motor symptoms after the length of the CAG repeat (bottom) had been taken into account were compared for the seven most abundant HD-mutation-bearing chromosome haplotypes, shown in Figure 3, across the HTT_Hap region (from Figure 1B).

(C) The distributions of expanded CAG repeat length (top), age at the onset of motor symptoms (middle), and residual of age at the onset of motor symptoms after the length of the CAG repeat (bottom) had been taken into account were compared for the 11 most common control haplotypes across the HTT_Hap region (see Figure 1B).

Each distribution is shown as a box plot, where the top, middle and bottom of the box represent the 75th percentile, median, and 25th percentile data points, respectively. The top and bottom horizontal lines (i.e., top and bottom whiskers, connected to the box by a vertical dashed line) represent the maximum and minimum data points, respectively, after outliers were removed by a standard interquartile method (1.5 × interquartile). p values of ANOVA models are shown in parentheses.

are associated with differences in local haplotype might have a modifier effect. To test this possibility, across HTT (21 SNPs) we defined 11 haplotypes that were most abundant on control chromosomes; each had a population frequency of >2%. Together, these 11 haplotypes accounted for ~79% of the normal chromosomes present in the 699 HD individuals genotyped. We modeled the residual of age at onset after accounting for the effect of the expanded CAG repeat as a function of normal chromosome some HTT haplotype in each individual and found no significant evidence of disease modification (Figure 2C).

As was the case for the HD-mutation-bearing HTT haplotypes, posthoc testing did not reveal any significant differences between individual haplotypes (data not shown). Thus, as with the HTT region of HD-mutation-bearing chromosomes, the HTT region of at least the majority of normal chromosomes does not act as a modifier of the age at the onset of motor symptoms in HD.
Underlying Mechanisms of Synthetic Association in the HD-Control GWAS

As noted above, the overall case-control association analysis was remarkable for yielding the most-significant association signals at a considerable physical distance from HTT. Because the causal mutation located in exon 1 of HTT is well established, our data provide a rare opportunity for investigating the underlying mechanisms of such synthetic association in an actual experimental dataset. It was evident from the haplotype analyses that the minor alleles for rs12641989 (A) in RGS12 and rs11248108 (A) in RNF4 (MIM 602850) were both overrepresented in the HD cohort. In the case of the latter marker, this overrepresentation was a result of its presence on the 165 major HD haplotype chromosomes with the 93 kb shared extended haplotype; in the case of the former marker, the overrepresentation was a result of its presence on the majority of the 330 HD chromosomes that bear HTT_Hap, 1, all of which also have the delta2642 deletion allele. However, none of the SNP markers in HTT itself yielded scores as highly significant as these distant markers. Clearly, the relative significance of the SNPs did not correlate with their physical distance from the causal variation (Figure S2A), but as expected, it did correlate strongly with the difference in allele frequency between cases and controls (Figure S2B). Although for these markers the absolute difference in minor-allele frequency between cases and controls was slightly correlated with distance from the expanded CAG (Figure S2C), we observed a much stronger correlation between control minor-allele frequency and association significance (Figure S2D). On the basis of these observations, we hypothesized that the synthetic association signals in the HD case-control GWAS were created by a combination of the existence of an extended version of an ancestral major haplotype represented by HTT_Hap1 (and probably HTT_Hap5) and the presence on chromosomes with that haplotype of minor alleles for some SNP markers that show low minor-allele frequency in controls.

To test this hypothesis, we investigated the relationship between control minor-allele frequencies and the significance of the association signals across the region. In order to discern patterns of recombination across the region as reflected in the major HD haplotype (Figure 4A), we used those SNPs whose minor alleles were significantly overrepresented in HD subjects as compared to controls (p < 0.05 in the original association analysis). We calculated the minor-allele frequency in controls (black) and the frequency of the same allele in HD subjects (red), in HD subjects with the minor (deletion) allele for delta2642 (green), and on HD-mutation-bearing chromosomes with the minor allele for delta2642 (blue). Grouping in this way provided a gradual enrichment for HD-mutation-bearing chromosomes with the major HD haplotype. For example, ~25.8% of all chromosomes in HD subjects (red), ~50% of all chromosomes in those HD subjects with a delta2642 deletion (green), and close to 100% of HD chromosomes that themselves had a delta2642 deletion (blue) represented this haplotype. As shown in Figure 4A, allele frequencies for all SNP minor alleles, except rs2857845, rs363082, and rs1730768 (arrows),
Genetic modifiers of Huntington’s disease

Supplementary Article 1

Figure 4. Effect of Allele Frequency in Controls and Recombination on Significance in Association Analysis

To test whether allele frequency in controls plays an important role in determining significance in association analysis, we analyzed data sets that were progressively enriched for the HD major haplotype.

A) For each marker, the allele frequency of that allele which is minor in the control population was plotted against its genomic location in the following groups: (1) all controls (1676 controls; black), (2) all cases (699 HD subjects; red), (3) cases with at least one delta2642 deletion allele (384 HD subjects; green), and (4) HD mutation-bearing chromosomes with a delta2642 deletion allele (562 chromosomes; blue). To simplify the analysis, we showed those SNPs whose minor allele was overrepresented in HD subjects versus controls in the original association analysis (p < 0.05; 41 SNPs). The boundaries of “nonrecombination” (blue) and “mild recombination” (red) zones represent arbitrary cut-offs (80% and 50%, respectively) based on allele frequency on HD-mutation-bearing chromosomes.

B) Association analyses were performed for all SNPs in the nonrecombination zone; each of the enrichment groups from (A) was compared to controls (allele test). The significance of the apparent HD risk alleles (overrepresented in HD compared to controls; p < 0.05 in the original dominant model association analysis involving all HD subjects and controls) (y axis) was plotted against the corresponding allele frequency in controls (x axis).

C) Association analyses were performed for all SNPs in the mild recombination zone; each of the enrichment groups from (A) was compared to controls (allele test). The significance of the apparent HD risk alleles (y axis) was plotted against the corresponding allele frequency in controls (x axis).

D) To eliminate the effect of recombination, we performed an association analysis (allele test) by comparing allele frequencies of all SNPs on the HD-mutation-bearing chromosomes with Ext_Hap_1 to their allele frequencies on control chromosomes. Significance (y axis) was plotted against frequency of apparent risk allele (x axis).

increased as the major HD haplotype was enriched (red to green to blue); this is consistent with our haplotyping results, which indicated that rs2857845, rs363082 and rs1730768 tag other haplotypes.

On the basis of the apparent decay in the major HD haplotype (i.e., Figure 4A; blue trace, excluding rs363082), we arbitrarily divided the chromosome 4p region into a relative “nonrecombination” zone (control minor allele present on >80% of HD major haplotype chromosomes; blue bar) and a relative “mild recombination” zone (control minor allele present on >50% but <80% of HD major-haplotype chromosomes; red bars). In both zones, we then tested SNP markers within each of the haplotype-enrichment groups separately for an association with HD. We tested the following groups: (1) all controls versus all cases (red); (2) all controls versus only those cases with a delta2642 deletion allele (green); and (3) all control chromosomes versus only those chromosomes bearing a HD mutation and having a delta2642 deletion allele (blue) (Figures 4B and 4C). In the nonrecombination zone (118 kb), we observed a strong negative correlation between the frequency of a SNP’s apparent risk allele (i.e., the allele increased in HD subjects) in controls and the significance of association with that SNP. The correlation became progressively stronger as the sample was enriched for the major HD haplotype (red to green to blue),
supporting the notion that the control allele frequency plays an important role in creating synthetic associations (Figure 4B). In the mild recombination zone (938 kb, excluding the 118 kb nonrecombination zone), we used only SNPs that showed nominally significant differences between HD subjects and controls in the original association analysis (dominant-model p value < 0.05) because inclusion of nonsignificant SNPs would add noise to the analysis and complicate interpretation. Similarly to the nonrecombination region findings, we observed a negative correlation between control minor-allele frequency and association significance, although the significance values were considerably lower than those achieved in the nonrecombination zone at a similar control allele frequency, suggesting a strong effect of recombinational decay of the major HD haplotype (Figure 4C). Taken together, these results imply that the control minor-allele frequency and the recombination frequency, rather than physical distance between the SNP and the causative defect, play important roles in determining synthetic associations.

To confirm this conclusion, we compared the subset of major-haplotype HD chromosomes that shared the entire extended haplotype of 938 kb from rs11248108 to rs12641989 to control chromosomes. This permitted us to compare the strength of association across SNPs without encountering any effect of recombination. As shown in Figure 4D, markers across this segment showed a smooth inverse relationship between the strength of association and allele frequency on control chromosomes, confirming the notion that control minor-allele frequency plays a critical role in synthetic associations. Together, our results strongly indicate that synthetic association signals in HD were contributed by a combination of a frequent, extended HD haplotype in cases and the inclusion of markers with relatively low minor-allele frequencies in the genotyping array of controls.

Discussion

The results of the case-control SNP association analysis of HD and the subsequent Htt-region haplotype analysis that we performed here, taken in historical context, have implications for applying the approach to chromosomal localization and identification of other strong-effect human genetic defects. The approach might be particularly useful for relatively rare phenotypes that result in small family sizes, which preclude effective linkage analysis, and for founder mutations that contribute only a subset of cases for a phenotype of diverse etiology. The genetic defect causing HD was identified almost two decades ago when our knowledge of human genetic variation was far less evolved and the technologies for performing human genetic studies were far less powerful than they are today. Its identification came after genetic linkage mapping, first with restriction-fragment length polymorphisms, and subsequently with PCR-based multi-allele markers, in HD-affected families confined the disease mutation to a region of 2.2 Mb in 4p16.3.\textsuperscript{1,20} This candidate region was then narrowed by association analysis, in which the biallelic markers presented the apparent quandary of closely spaced polymorphisms that alternatively showed either strong or no evidence for association. The conundrum was explained by the power of the multi-allele markers to resolve multiple ancestral haplotypes, a particularly frequent one of which was marked by the minor deletion allele of a 3 bp in/del polymorphism (delta2642) that ultimately proved to represent a codon in Htt.\textsuperscript{16,39,22} The comparison of markers on HD chromosomes bearing this minor allele focused the search for the genetic defect to a 150 kb segment with the Htt CAG repeat and led to its identification.

Not unexpectedly, the data generated from our current analysis indicate that a case-control genome-wide SNP-association approach could have quickly accomplished the initial mapping of the HD genetic defect to 4p16.3 and that it could have done so with as few as 50 unrelated HD cases and 1,600 controls, despite the presence of multiple founder chromosomes. However, the analysis of individual SNPs would not have finely localized the CAG mutation; the most significantly associated marker is in Rgs12, 343 kb from the defect. It should also be noted that application of current whole-exome or even whole-genome next-generation sequencing strategies would not have identified the HD mutation because sizing even longer normal alleles is problematic with short-read technologies. Thus, like the multi-allele-marker haplotype analysis that actually led to the identification of the expanded CAG repeat, SNP-based haplotype analysis would have been required to home in on the defect after the initial association analysis placed it in region 4p16.3. rs12641989 yielded the strongest association signal because it was present on the most frequent ancestral HD haplotype, it had a low minor-allele frequency in the control population, there was a low recombination rate across this region, and this low recombination rate allowed an extended shared haplotype. Our data strongly suggest that SNP arrays aimed at genotyping markers with even lower control minor-allele frequencies would be yet more powerful for initial chromosomal localization of founder mutations. However, because a defect localized by association with a rare allele might be at a considerable physical distance from the marker, subsequent identification of the actual defect could benefit from detailed haplotype analysis that takes into account the local recombination landscape. The same argument can be made for phenotypic associations linked to genomic segments introduced into a population by admixture or to phenotypes where only a minor proportion of cases are due to a founder mutation. Indeed, case-control simulations indicate that for a SNP with a control minor-allele frequency of 0.01, a founder haplotype bearing such an allele need be present in less than 60 cases in 1,000 or 100 cases in 5,000 to yield genome-wide significance (p < 5 x 10^-5; Figure S3).
Our SNP haplotype analysis of the HTT region, which extends previous lower-resolution investigations, is revealing with regard to the structure of the HD population. The studies that led to the identification of the CAG expansion mutation investigated only 78 HD chromosomes without regard to population ancestry and identified 26 different haplotypes, although these involved markers widely dispersed over the large 2.5 Mb HD candidate region then defined by recombination events in families. There was far less diversity in the immediate region of the mutation: Approximately one-third of chromosomes shared a common set of alleles, including the deletion allele at HTT codon 2642. More recently, Warby et al. identified 22 HTT SNPs that permitted the definition of a haplogroup representing 95% of 268 HD European chromosomes. However, their analysis allowed for diversity at many additional intervening SNPs within HTT, indicating that this single haplogroup combines multiple HTT haplotypes. Though the analysis here was also restricted to individuals of European ancestry, the delineation of multiple dense SNP haplotypes spanning the HD mutation is most consistent with multiple independent origins of the HD-mutation-bearing chromosomes in this population.

The seven most abundant HTT-region haplotypes constituted 582 of the 699 chromosomes, whereas the remaining 117 HD-mutation-bearing chromosomes accounted for 48 additional haplotypes (HTT_Hap_Other). Although many of these haplotypes, such as HTT_Hap_1 and HTT_Hap_5, might be derived from one another by historical crossovers, it seems likely that there have been a minimum of four founder events given that there are HD-mutation-bearing chromosomes representing all four two-marker haplotypes for SNPs rs2795296 and rs3856973, which flank the CAG repeat (A-A, four in HTT_Hap_Other; A-G, HTT_Hap_3 and ten in HTT_Hap_Other; G-A, HTT_Hap_4 and three in HTT_Hap_Other; G-G, HTT_Hap_1, HTT_Hap_2, HTT_Hap_5, HTT_Hap_6, HTT_Hap_7, and 31 in HTT_Hap_Other).

Despite this diversity, one founder event accounts for about half of all these European HD chromosomes (HTT_Hap_1 and probably the derivative HTT_Hap_5). It corresponds with the most frequent HD haplotype defined in earlier studies given that it is marked by the delta2642 deletion allele. Of the most frequent HD haplotypes, HTT_Hap_2 and HTT_Hap_3 show little or no increase in frequency relative to control chromosomes, whereas HTT_Hap_1 shows an increase of more than 5-fold. Interestingly, almost 50% of the HTT_Hap_1 HD-mutation-bearing chromosomes share an identical SNP haplotype that reaches far beyond the confines of HTT and extends over at least 938 kb, reflecting the relatively low recombination rate across this segment and reinforcing the view that many of the other observed haplotypes represent independently originating HD mutations. This pattern also fits predictions that are based on the analysis of reproductive success in HD.22

Delineation of the haplotype structure of the 4p16.3 segment has permitted us to test for cis-acting modifiers of HD pathogenesis. For the seven most abundant HTT haplotypes, we observed no significant difference in the distribution of expanded CAG repeats, indicating that at least for the substantial majority of HD chromosomes, there is no evident effect of local variation on the instability of the expanded CAG repeat. Previous analysis suggested enrichment of a particular haplogroup on both HD chromosomes and control chromosomes with CAG repeats high in the normal range, raising the possibility that this haplogroup is predisposed to instability.13 The low frequency of this haplogroup among East Asians has been proposed as the basis for the lower prevalence of HD in this population relative to Europeans.23 Our data are not directly comparable to the haplogroup analyses, but the lack of an effect of the local haplotype on instability within the expanded CAG allele range does not preclude an effect within the high normal HD range. Similarly, although we cannot speak to the CAG allele distribution on the control chromosomes used here, it is likely that the HD_Hap_1 haplotype in the normal population is associated with higher-range normal alleles that act as a reservoir of new mutations to HD because such high normal alleles might be more likely to expand into the HD range. Indeed, we have previously noted that the major haplotype defined by the delta2642 deletion allele marks chromosomes that gave rise to sporadic cases of HD in “new mutation” families, indicating that the current pool of HTT_Hap_1 HD cases is descended from multiple normal individuals who transmitted a newly expanded allele on this haplotype.24

The detailed haplotype analysis also provided the basis for testing the hypothesis that local genetic variation modifies HD pathogenesis through either a cis- or a trans-acting effect on HTT expression. Neither the extended shared haplotype nor the seven most abundant HTT haplotypes on HD-mutation-bearing chromosomes showed evidence of influencing age at the onset of motor symptoms after the effect of the expanded CAG repeat length, which is the main driver of pathogenesis, was taken into account. Similarly, the HTT haplotypes on the normal chromosomes present in HD subjects showed no evidence of an impact on the age at onset, consistent with the previous finding that neither the length of the normal CAG allele nor the presence of a second expanded CAG allele has a significant impact on the rate of pathogenesis leading to the onset of motor symptoms.7 Although it is conceivable that the local HTT haplotype might still influence some other HD phenotype, the age at the onset of motor symptoms correlates strongly with the length of the expanded CAG allele and shows evidence of being altered by one or more modifier genes.

Our study also does not exclude the possibility that one or more rare haplotypes among the HD_Hap_Other grouping does involve a cis-modifier effect on either CAG repeat instability or the age at the onset of motor
Symptoms or that a rare <2% frequency control haplotype might exert a \textit{trans} effect. However, our findings do indicate that no such effects occur in most HD subjects, and they suggest instead that the emphasis should be on identifying \textit{trans}-acting modifiers located elsewhere in the genome. The ability to attribute the full effect of the \textit{HTT} locus to the length of the expanded CAG repeat, without having to consider other genetic variation at the locus, will facilitate the analysis of genome-wide association scans aimed at identifying such modifier genes; such analysis is a high priority in HD because valid targets for therapeutic intervention are badly needed.

**Supplemental Data**

Supplemental Data include three figures and three tables and can be found with this article online at \url{http://www.cell.com/AHFG}.

**Acknowledgments**

We thank the HD research participants and their families, who made this work possible; the Harvard Brain Tissue Resource Center; the COHORT Study of the Huntington Study Group; and the Myocardial Infarction Genetics Consortium (MiGen) study, which was funded by grants from the U.S. National Institutes of Health and National Heart, Lung, and Blood Institute's STAMPED genomics research program (RO1 HL087676) and the National Center for Research Resources (U54 RR020278). We also thank investigators and contributors to the HD-MAPS study: Alexandra Durr, Adam Rosenblatt, Luigi Frati, Susan Perlman, P. Michael Conneally, Mary Lou Klimek, Melissa Diggin, Tiffany Hadzi, and Ayana Duckett. This work was supported by grants from the National Institute of Neurological Disorders and Stroke (NS16367) (Huntington’s Disease Center Without Walls) and NS32765, the CHDI Foundation, and the Huntington’s Disease Society of America’s Coalition for the Cure. E.M.R. received a scholarship from the Fundação para a Ciência e a Tecnologia de Portugal (SFRH/BD/44335/2008).

Received: November 23, 2011  
Revised: December 24, 2011  
Accepted: January 9, 2012  
Published online: March 1, 2012

**Web Resources**

The URLs for data presented herein are as follows:

- International HapMap Project, \url{http://hapmap.ncbi.nlm.nih.gov/}
- MACH Home Page, \url{http://www.sph.umich.edu/csg/abecasis/MACH/index.html}
- Online Mendelian Inheritance in Man (OMIM), \url{http://www.omim.org/}

**References**


Supplementary Article 2: TAA repeat variation in the GRIK2 gene does not influence age at onset in Huntington's disease.


Reprinted with the permission of Elsevier
License number 3190381064259
TAA repeat variation in the GRIK2 gene does not influence age at onset in Huntington’s disease

Ji-Hyun Lee a, b, *, Jong-Min Lee a, Eilana Marisa Ramos a, c, Tammy Gillis d, Jayalakshmi S. Mysore e, Shotaro Kishikawa a, Tiffany Hadzi d, Audrey E. Hendricks e, Michael R. Hayden f, Patrick J. Morrison g, h, Martha Nance i, Christopher A. Ross i, Russell L. Margolis j, Ferdinando Squitieri k, Cinzia Gellera k, Estrella Gomez-Tortosa a, Carmen Ayuso a, Oksana Suchowersky l, Ronald J. Trent m, Elizabeth McCusker n, Andrea Novelletto o, Marina Frontali o, Randi Jones o, Tetsuo Ashizawa p, Samuel Frank q, Marie-Helene Saint-Hilaire q, Steven M. Hersch y, Herminia D. Rosas y, Diane Lucente y, Madaline B. Harrison y, Andrea Zanko y, Ruth K. Abramson y, Karen Marder z, Jorge Sequeiros c, a, b, a, b, G. Bernhard Landwehrmeyer a, b, On behalf of the Registry Study of the European Huntington’s Disease Network, Ira Shoulson a, c, On behalf of the Huntington Study Group COHORT project, Richard H. Myers a, Marcy E. MacDonald a, James F. Gusella a, b, d

* Corresponding author at: Department of Pharmacology, Pharmacogenomic Research Center for Membrane Transporters and Research Center for Human Natural Defense System, Yonsei University College of Medicine, 50 Yonsei-ro, Soodoam-gu, Seoul 120-752, Korea. Fax: +82 2 313 1894.
E-mail address: jhyun@yuhs.ac (J.-H. Lee).

0006-291X/ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymbcr.2012.06.120

ABSTRACT

Huntington’s disease is a neurodegenerative disorder caused by an expanded CAG trinucleotide repeat whose length is the major determinant of age at onset but remaining variation appears to be due in part

ARTICLE INFO

Article history:
Received 16 June 2012
Available online 3 July 2012

218 Genetic modifiers of Huntington’s disease
1. Introduction

Huntington’s disease (HD) is an autosomal dominant, progressive neurodegenerative disorder that presents with motor dysfunction, cognitive decline and psychiatric disturbances due to expansion of a trinucleotide CAG repeat encoding a polyglutamine tract in the huntingtin protein [1,2]. There is a strong inverse correlation between age at diagnosis by onset of motor signs and the CAG repeat length, which accounts for up to 67% of the overall variance [3–5]. The remaining variation is strongly heritable, indicating a contribution of modifier genes to determining age at onset [6–8].

GRK2 (Glutamate receptor, ionotropic kainate 2), encoding the GluR6 subunit of the predominant excitatory neurotransmitter receptor family in the human brain, is an attractive candidate as an HD modifier because of its potential role in excitotoxic cell death [9]. A TAA trinucleotide repeat polymorphism in the 3’untranslated region (3’UTR) of GRK2 mRNA was reported to be associated with age at onset of diagnostic motor signs in small studies from the United Kingdom, New England, France, India, and Italy [10–14], the largest of which comprised less than 300 subjects. The two initial studies suggested that the modifier effect was due to genetic variation on chromosomes carrying a 16 TAA repeat allele. The genetic variation responsible for the earlier than expected age at onset was implicated as the TAA repeat itself by subsequent haplotype analysis [15]. However, the mechanism by which this polymorphism could act is not certain and might conceivably include effects on alternative splicing, editing, stability or translational regulation of the GRK2 mRNA. Consequently, before embarking on extensive molecular analyses to define the mechanism of action, we carried out a comprehensive analysis of a much larger set of HD subjects than has previously been examined to confirm GRK2 as a genetic modifier of HD pathogenesis.

2. Methods and materials

2.1. HD samples

HD patient DNA samples from individuals with known age at motor, psychiatric and/or cognitive onset were obtained from ongoing genetic studies at the MGH HD Center Without Walls, members of the HD-MAPS collaboration, post-mortem brain specimens (Harvard Brain Tissue Resource Center and the UCLA Brain Bank), and two large observational studies: the Huntington Study Group’s COHORT project and the European Huntington Disease Network’s REGISTRY study. In total, 2,911 HD heterozygote subjects with one expanded HD allele and known age of motor onset (2,362 individuals), psychiatric onset (547 individuals) and/or cognitive onset (210 individuals) were genotyped for the GRK2 TAA repeat polymorphism and for the HD CAG repeat as described previously [9,16]. The mean age at onset was 42.6 (range, 4–92) and the mean expanded allele HD CAG repeat length was 45.1 (range, 36–98). This study used de-identified DNA samples and was approved by the Institutional Review Board of the Partners HealthCare System.

2.2. Data analysis

This study utilized samples from 2,911 HD heterozygote subjects with one expanded HD allele and known age of motor onset (2,365 individuals), psychiatric onset (547 individuals) and/or cognitive onset (210 individuals). Potential modifier effects were explored by adding the factor to a linear regression model relating the natural log-transformed age at onset to HD CAG repeat length and determining the degree of improvement in goodness-of-fit. Since the regression plots for onset of symptoms in the three different domains show quite different relationships between natural log-transformed age at onset and HD CAG repeat length, each category of age at onset was analyzed separately. SPSS 11.5 (SPSS Inc., Chicago, IL, USA) was used for all statistical analyses.

3. Results

The GRK2 TAA repeat displays 7 alleles, ranging from 12 to 17 TAAs. The 16 TAA allele was reported previously to be associated with earlier age of onset in HD [10,11]. We initially tested potential modifier effects by determining whether adding the GRK2 genotype (dominant model) had a significant impact on a linear regression model relating the natural log-transformed age at onset to HD CAG repeat length. Separate tests were performed for 1) presence of at least one 16 TAA allele, 2) presence of at least one of the two longest alleles (16 or 17 TAAs), and 3) the GRK2 TAA repeat as a continuous trait, based upon the larger of the two alleles present in each individual. Fig. 1A shows the relationship in 2,362 HD subjects of age at onset of diagnostic motor signs with CAG repeat length and Fig. 1B reveals a lack of any significant impact of adding GRK2 to the model on improving the R² value.

To avoid the potential disproportionate impact of subjects with extreme CAG lengths reported by Lee et al. [12], we calculated standardized residuals for the 2,205 subjects with CAG repeats in the 40–53 range, relative to the regression line generated in that report. This analysis (Fig. 1E–D) again shows no evidence of an effect of GRK2 genotype on motor onset. We also performed a number of analyses in which we subdivided these samples into HD subjects with extremes of age at onset associated with each CAG repeat length using different cut-offs for inclusion or exclusion and in no case did comparison of GRK2 genotypes between the two extreme groups reveal a significant difference. Thus, our findings in this large dataset indicate that the GRK2 TAA repeat polymorphism, previously thought to be an HD modifier, is not significantly associated with deviation in HD age of motor onset from that expected based upon CAG repeat length.
Although HD is typically diagnosed based upon onset of motor signs, the age at onset for cognitive and/or psychiatric symptoms is available in a subset of this sample. Directly comparable to Fig. 1, Fig. 2 shows no significant effect of the GRK2 TAA repeat polymorphism on age at onset of either psychiatric (Fig. 2A–B; 547 subjects) or cognitive (Fig. 2C–D; 210 subjects) symptoms.

4. Discussion

It is well established that the expanded CAG repeat in the HD gene is the primary determinant of age of onset of clinical symptoms, but that the mutation alone does not explain all of the variation observed. The remaining variance displays a high degree of heritability, supporting the existence of genetic modifiers, factors whose polymorphic variation influences the course of HD. Consequently, identification of genetic modifiers from human patients represents a potential route to discovering validated targets whose modulation would be expected to alter HD pathogenesis. To date, candidate gene strategies have produced a number of reports of potential modifiers from pathways postulated to be involved in HD pathogenesis [7,12,18–21], including several in which the GRK2 TAA repeat polymorphism was associated with altered onset age [10,11,15]. GRK2 encodes a protein integrally involved in mediating excitatory neurotransmission in the brain and its status as a genetic modifier would bolster the excitotoxicity model of HD. However, detailed investigation of its mode of action would entail a major investment of labor and research funds.

As genetic analysis of functional variants has argued strongly against other attractive modifier candidates, such as BDNF which showed no effect on age at onset in two large studies [22,23], we felt it important to confirm the GRK2 modifier effect in a much larger study sample than was previously tested. The increased power afforded by the much larger sample size and more comprehensive analytical methods in our current study argue strongly that GRK2 has no significant effect on HD pathogenesis leading to age of onset of motor symptoms. Albeit with smaller sample sizes, our study also does not support modifier effects on age of onset psychiatric or cognitive symptoms. These findings argue against the pursuit of detailed molecular and biological analysis of any putative functional effect of this GRK2 polymorphism as a clue to how to alter HD pathogenesis. In this regard, they are consistent with negative findings by Metzger et al. [24] in a German HD population and by Andresen et al. [25] in the Venezuela HD population, although the latter had relatively few 16 TAA alleles.

Our findings present a cautionary note for HD modifier studies in general since an apparent GRK2 effect was noted previously for motor onset in multiple studies [10–15]. These apparent positive associations may have been due to small, possibly unrepresentative samples and may have been disproportionately influenced by genotypic or phenotypic outliers.

In conclusion, our study results showed that the previously studied TAA repeat variation in the GRK2 gene does not influence age at onset in Huntington’s disease using the largest dataset assembled to date and suggest that similarly large-scale investiga-
Fig. 2. GRK2 modifier analysis in 547 HD subjects with known age at psychiatric onset and in 210 HD subjects with known age at cognitive onset. For the GRK2/TA repeat polymorphism, subjects with at least one 16 or 17 TA allele are displayed as closed orange and red circles, respectively, with all other subjects being represented by open black circles. (A) The relationship between HD expanded allele and log transformation of age at onset of psychiatric signs. (B) Summary statistics for linear regression analysis for data shown in (A) in each GRK2 allele mode. The near-significant result for analysis of the TA repeat as a continuous trait is skewed by the small number of extreme CAG repeat samples, as identified analysis of the bulk of subjects with CAG repeats of 40–53 yielded P = 0.503. (C) The relationship between HD expanded allele and log transformation of age at onset of cognitive signs. (D) Summary statistics for linear regression analysis for data shown in (C) in each GRK2 allele mode. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6. Authors’ contributions

JHL, participated in the design of the study, analyzed the data statistically and drafted the manuscript. JML assisted in statistical analysis of the data and critical revision of the manuscript. EMR, TG, JSM, and SK generated the molecular data while TH, AEH, MRH, PM, MN, CAR, RLM, FS, CG, EGT, CA, OS, RJT, EM, AN, MF, RJ, TA, SF, MPH, SMH, HDR, DI, MBH, AZ, RKA, KM, JS, CBI, and IS participated in design of the study and helped to generate the clinical data. RHM, MEM, and JFG conceived the study, participated in its design and critically revised the manuscript. All authors have read and approved the final manuscript.

Acknowledgments

The authors thank contributors to the HD-MAPS, COHORT and Registry studies, listed in Appendix I. This work was supported by grants from the National Institutes of Health NINDS Huntington’s Disease Center Without Walls NS16367, the CHDI Foundation Inc., and the Huntington’s Disease Society of America’s Coalition for the Cure. JHL received support from a National Research Foundation of Korea Grant funded by the Korean Government [NRF-2009-352-E00010]. EMR is the recipient of a scholarship from FCT (SFRH/BD/44335/2008). The Huntington Study Group COHORT project was supported by the CHDI Foundation, Inc. and the REGISTRY study of the European Huntington’s Disease Network was supported by the High Q Foundation.

References


Acknowledgments
Though only my name appears on the cover, many have helped and contributed, in various different ways, to my PhD dissertation. Therefore, these last lines are dedicated to all of those who made this possible.

I would like to start by thanking my supervisors, Drs. Isabel Alonso, Marcy MacDonald and Prof. Dr. Jorge Sequeiros, for giving me the opportunity and privilege of doing my PhD with them. Thank you for all the encouragement and trust placed in me and for always pushing me to go further. I would also like to thank Dr. James Gusella for always being available with his insightful comments, great advices and constant support.

A special thanks to my Chefinha for her immeasurable patience and enthusiasm not only as a mentor but also as a friend; for letting me follow my dreams all the way to the other side of the Ocean; and for always coming to my rescue via Gchat or Skype. I will never forget our astonishing adventure in the Aussieland, even though I didn’t get to (yet) surf with the Aussies and climb Sydney Harbour Bridge. [P.S. - I have to thank the European HD Network for organizing the World Congress of HD in Australia, giving me the perfect ‘excuse’ to finally realize my biggest dream.]

I would like to thank my colleagues at IBMC and CHGR for their various forms of support and help. Thank you to all my current and former lab mates at UnIGENe and MacDonald lab for the remarkable teamwork and for a friendly, energizing and fun place to work. Thank you also to all the coauthors of the presented papers for their excellent collaboration, in particular Dr. Roy Perlis and Dr. John Landers for giving me the opportunity to enlarge the scope of my research to other brain disorders.

I have to give a very special mention to Carolina for being my statistical adviser, my personal assistant and reviewer and more importantly a really good friend… and also to Martina and Kawtherina for being the big sisters I’ve never had and for showing me that it’s never too late to fight for what we want.

I was fortunate enough to have met an outstanding group of people that helped me adjust to a new country and whom I’m proud to call now friends. A big shout-out to Cláudio, Joana, Rosa, Jeff, Steffen, Indra and Matthew, my Bostonian bffs. Thank you for all those fun nights playing games, for our amazing camping/snowboard trips, for bearing my sports addiction. They have tried hard to help me stay sane through these last four years, a quite difficult task I should say. I greatly value their friendship and I deeply appreciate their belief in me. Thank you also to all my roommates for always making me
feel at home. [P.S.- I have to thank Mr. Tom Brady and the rest of the Patriots (and the Celtics and the Bruins) for giving me an excuse every week to go to the Tavern and not think about work for a few hours and mostly for being the best team a fan can ask for.]

I can’t forget the amazing group of friends I have back home. We may be separated by too many miles and only see each other once a year but true friends don’t fall apart. Thank you for not letting the distance overcome our friendship. And thank you for always welcoming me, as if I’ve never left, making every trip back home memorable.

It would be unfair to finish without thanking my family that, in spite of the long distance, has always been able to show their love and support. They’ve always found a way to make me feel at home despite the 5000km in between us - through postcards, funny homemade B’day videos and hilarious video calls during the holidays. I would like to specially thank my parents to whom I owe everything. For always letting me follow my dreams no matter how crazy they may have seemed or how far from home they may have led me.

And finally, I would like to thank with all my heart the most important person in my life, my lil’ sister, well, for being my sister, my confidant, my best friend; for always being there for me, through the best and the worst, at a distance of a phone/video call; for pushing me to dream even higher; and for believing in me more than I believe in myself sometimes.

“You’re my sweetheeeeaart”.

Thank you all. I wouldn’t have come this far without any of you.
Acknowledgments

Institutions:

Porto, Portugal

Boston, USA

Funding:

Eliana Marisa Ramos is the recipient of a scholarship from FCT (SFRH/BD/44335/2008).
The work presented in this dissertation was also funded by NIH, CHDI Foundation and HDSA grants.