Replication of twelve association studies for Huntington's disease residual age of onset in large Venezuelan kindreds

J M Andresen, J Gayán, S S Cherny, D Brocklebank, G Alkorta-Aranburu, E A Addis, The US-Venezuela Collaborative Research Group, L R Cardon, D E Housman and N S Wexler

J. Med. Genet. 2007;44;44-50; originally published online 3 Oct 2006; doi:10.1136/jmg.2006.045153

ORIGINAL ARTICLE

Replication of twelve association studies for Huntington's disease residual age of onset in large Venezuelan kindreds

J M Andresen, J Gayán, S S Cherny, D Brocklebank, G Alkorta-Aranburu, E A Addis, The US-Venezuela Collaborative Research Group, L R Cardon, D E Housman, N S Wexler

J Med Genet 2007;**44**:44–50. doi: 10.1136/jmg.2006.045153

See end of article for authors' affiliations

Correspondence to: Dr J M Andresen, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; jma@mit.edu

Received 29 June 2006
Revised 26 August 2006
Accepted 1 September 2006
Published Online First 13 September 2006

Background: The major determinant of age of onset in Huntington's disease is the length of the causative triplet CAG repeat. Significant variance remains, however, in residual age of onset even after repeat length is factored out. Many genetic polymorphisms have previously shown evidence of association with age of onset of Huntington's disease in several different populations.
Objective: To replicate these genetic association tests in 443 affected people from a large set of kindreds from Venezuela.
Methods: Previously tested polymorphisms were analysed in the *HD* gene itself (*HD*), the GluR6 kainate glutamate receptor (*GRIK2*), apolipoprotein E (*APOE*), the transcriptional coactivator CA150 (*TCERG1*), the ubiquitin carboxy-terminal hydrolase L1 (*UCHL1*), p53 (*TP53*), caspase-activated DNase (*DFFB*), and the NR2A and NR2B glutamate receptor subunits (*GRIN2A*, *GRIN2B*).
Results: The *GRIN2A* single-nucleotide polymorphism explains a small but considerable amount of additional variance in residual age of onset in our sample. The *TCERG1* microsatellite shows a trend towards association but does not reach statistical significance, perhaps because of the uninformative nature of the polymorphism caused by extreme allele frequencies. We did not replicate the genetic association of any of the other genes.
Conclusions: *GRIN2A* and *TCERG1* may show true association with residual age of onset for Huntington's disease. The most surprising negative result is for the *GRIK2* $(TAA)_n$ polymorphism, which has previously shown association with age of onset in four independent populations with Huntington's disease. The lack of association in the Venezuelan kindreds may be due to the extremely low frequency of the key $(TAA)_{16}$ allele in this population.

Huntington's disease (OMIM 143100) is a devastating neurodegenerative disease that causes involuntary movements, personality changes, severe emotional disturbance and cognitive decline.[1] This autosomal dominant disorder usually strikes in the third or fourth decade of life and leads inexorably to death 10–25 years after onset. Current treatments address only symptoms; none slows down the Huntington's disease progressive neurodegeneration.

Huntington's disease is an autosomal dominant disorder caused by the expansion of an unstable CAG repeat embedded in the first exon of the *HD* gene, which leads to an expanded polyglutamine repeat in the huntingtin protein.[2] Normal alleles contain <35 CAG repeats. Alleles with 35–39 CAG repeats show incomplete penetrance. Expanded alleles with ⩾40 CAG repeats are fully penetrant in a normal human lifespan.

The age of onset varies inversely with the number of CAG repeats in the gene, and repeat length alone explains about 70% of the variance in age of onset.[2–9] The onset age of people with the same repeat length varies dramatically, however, which indicates that other factors must also contribute. Two large family-based studies have shown that genetic factors are probably the primary determinants of this residual variance in age of onset for Huntington's disease, assuming that shared environmental factors are negligible. In a group of sibling pairs from North America, Europe and Australia, 65–71% of the variance in age of onset depends on CAG repeat length alone, whereas 11–19% depends on factors shared by the siblings.[10 11] In our Venezuelan kindreds, approximately 72% of the variance in age of onset depends on CAG repeat length, whereas approximately 17% depends on additional genetic factors.[12]

This has led to a search for genetic factors that influence the age of onset. Association of the age of onset with polymorphisms has been reported previously in nine genes: the *HD* gene itself (*HD*),[13–15] the GluR6 kainate glutamate receptor (*GRIK2*),[6 15–17] apolipoprotein E (*APOE*),[18 19] the transcriptional coactivator CA150 (*TCERG1*),[20] the ubiquitin carboxy-terminal hydrolase L1 (*UCHL1*),[17] p53 (*TP53*),[21] caspase-activated DNase (*DFFB*),[21] and subunits of the NR2A and NR2B glutamate receptors (*GRIN2A*, *GRIN2B*).[22] We have repeated these association studies in the Venezuelan kindreds with Huntington's disease, genotyping 443 affected people and 361 unaffected family members.

MATERIALS AND METHODS

Choice of polymorphisms for association testing

The 12 polymorphisms in the nine genes shown in Table 1 were, to the best of our knowledge, a comprehensive list of association studies reporting a p value <0.05 when we undertook our task of replication.

Single-nucleotide polymorphism genotyping

The region surrounding each single-nucleotide polymorphism (SNP) was amplified with primers selected using the default settings of Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Polymerase chain reactions (PCRs) were carried out with 25 ng of genomic DNA in 20 mM $(NH_4)_2SO_4$, 75 mM Tris, pH 8.8, 1.5 mM $MgSO_4$, 0.17 mM of each deoxynucleotide triphosphate, 0.5 μM each of forward and reverse primers (Invitrogen, California, USA) and 1 U Taq polymerase using a touchdown programme (40 cycles of 30 s at 94℃; 30 s at 65℃ minus 0.5℃/cycle; 1 min at 72℃; then 10

Abbreviations: PCR, polymerase chain reaction; QTDT, Quantitative Transmission Disequilibrium Test; SNP, single-nucleotide polymorphism

cycles of 30 s at 94℃; 30 s at 55℃; and 1 min at 72℃) in a PTC-225 Pelthier Thermal Cycler. PCR product (25 µl) was denatured by adding 180 µl of denaturing solution: 20 mg/ml NaOH, 2 M NaCl and 25 mM EDTA. Half (100 µl) of the denatured PCR product was spotted to each of two positively charged Hybond-N membranes (Amersham Pharmacia Biotech, New Jersey, USA). After denaturation, the filters were placed on Whatman paper soaked in 2× standard sodium citrate for at least 2 min to neutralise and then left to dry. Prehybridisation was carried out at 50℃ in hybridisation buffer (3M tetramethyl ammonium chloride, 5× Denhardt's solution, 1 mM EDTA, 10 mM sodium phosphate, pH 6.8, 0.5% lauryl sulphate, 0.02 mg/ml yeast RNA) with 60 nM unlabelled competitor allele-specific oligonucleotide. The alternative allele-specific oligonucleotide was end-labelled with [γ^{33}P] adenosine triphosphate using T4 polynucleotide kinase (New England Biolabs, Massachusetts, USA). The labelled probe was added to the hybridisation mix to a final concentration of 5 nmol/l and incubated at 50℃ overnight. Membranes were washed three times in 3 M tetramethyl ammonium chloride, 2 M NaCl and 25 mM EDTA for 20–30 min in the hybridisation bottles at 50℃. The filters were exposed to Phosphorimager screens overnight and read in a STORM 850 Scanner (Molecular Dynamics, New Jersey, USA).

Microsatellite genotyping

We performed polymerase chain reactions as described for the *GRIK2* $(TAA)_n$ polymorphism,[23] the *TCERG1* $(Gln\text{-}Ala)_n$ polymorphism[20] and the *HD* $(CCG)_n$ polymorphism.[24] Fragment lengths were determined using capillary electrophoresis on an Applied Biosystems 3700 DNA Analyzer. As the estimated fragment length from a run on an Applied Biosystems machine can differ from the true fragment length by several base pairs, precise repeat numbers of five to ten samples were identified by DNA sequencing for the *GRIK2* TAA repeat to set the phase of the capillary electrophoresis reads. This small source of error may also have affected the run length of *TCERG1* allele, where our major allele ran at 306 base pairs (bp) compared with 304 bp in the previous study.[20] This error can cause differences for samples run on different machines, but does not fluctuate for samples run on a single machine. We believe that both the 306 bp fragment we measure and the 304 bp fragment the prior study measured correspond to the imperfect $(Gln\text{-}Ala)_{38}$ repeat (actually consisting of 34 Gln-Ala dipeptides, with four Gln-Val dipeptides interspersed) in the reference sequence for the gene (NM_001040006). Both studies used the same primers for PCR to generate an expected fragment size of 306 bp for a 38-repeat allele.

The Venezuelan kindreds

These large kindreds from Venezuela with a high incidence of Huntington's disease have been described previously.[12] For the present study, we used a subset of 794 people (affected people and relatives) who have been genotyped. Of these, 443 subjects who are heterozygotes for the *HD* gene (only one allele with ⩾35 CAG repeats) and have been diagnosed with Huntington's disease were used in the association statistical analyses. Pedigree statistics and Mendelian inheritance were checked using the program Pedstats.[25]

Statistical analyses

The relationship between age of onset of Huntington's disease and CAG repeat length is curvilinear. To model this relationship, we fitted a simple linear regression predicting the natural log of age of onset of Huntington's disease from CAG repeat length. To evaluate the effect of genetic markers in candidate genes on age of onset of Huntington's disease, we included these markers as predictors in the regression model. For biallelic markers, a linear allelic or genotypic predictor was used, unless evidence for a more complex model was observed. For multiallelic markers, a categorical predictor was created from either allelic or genotypic classes, and other models were tested if evidence supported them. It is important to note here that the observations used in these regression analyses are not independent because many of these subjects are members of the same families and are therefore related. This may bias the results.

We also used linear regression to compute a residual age of onset, after controlling for CAG repeat length, for subsequent genetic analyses in the software package QTDT (Quantitative Transmission Disequilibrium Test).[26] QTDT has implemented several models of association, such as the total association test, modelling all the evidence for association, and the orthogonal test, which fits between and within families effects, thus controlling for potential stratification. The orthogonal test should be used in a variance components framework to control for environmental and background polygenic effects, and allows for the analysis of association in extended pedigrees such as the Venezuelan kindreds.

Finally, differences in mean age of onset for the different allelic or genotypic classes were assessed by t test.

We have performed multiple statistical tests in several genetic markers, but are reporting p values that are uncorrected for this multiple testing. The purpose of this study was to report evidence for association in an independent sample as a possible replication of several published candidate gene associations.

RESULTS

Association with *HD*

The CAG repeat that causes Huntington's disease explains around 70% of the variance in age of onset.[2–9] Two additional polymorphisms in the *HD* gene that change the primary structure of the huntington protein have been tested for genetic association: a $(CCG)_n$ repeat length polymorphism in exon 1 and deletion of a glutamic acid residue at residue 2642 in exon 58 (table 1).

A study of 84 French patients showed association between age of onset and the *HD* Δ2642 polymorphism ($p<0.05$).[14] This correlation was not confirmed in studies of 293 patients from the UK,[6] 126 patients from Italy,[27] or 518 patients from North America, Europe and Australia[28] (table 1).

We also fail to replicate this result in the Venezuelan sample (table 2). The Δ2642 polymorphism explains no additional variance in age of onset, with heterozygotes having an average age of onset almost identical to homozygotes.

A study of 77 Eastern Indian patients showed association between age of onset and the *HD* $(CCG)_n$ polymorphism that codes for a polyproline repeat immediately after the polyglutamine repeat ($p = 0.011$).[15] This correlation was not confirmed in the French study of 84 patients,[14] the Italian study of 126 patients,[27] a Taiwanese study of 38 patients[30] or a Russian study of 57 patients[29] (table 1).

This correlation is also not confirmed in the Venezuelan sample (table 2). Homozygotes for the most common $(CCG)_7$ allele have slightly higher ages of onset than most other genotype classes, but neither this effect nor the slight increase in R^2 (from 0.747 to 0.750) is statistically significant. Homozygotes for the $(CCG)_6$ allele have the highest average age of onset, but there are only two of them.

The normal, unexpanded *HD* allele has also been shown to have some association with age of onset in 754 patients from North America, Europe and Australia ($p = 0.012$), with longer repeat lengths in the normal *HD* gene being associated with later age of onset.[31] A study of 138 patients from Wales found a

Table 1 Previously reported genetic association tests of residual age of onset for Huntington's disease

Candidate gene	Polymorphism	Number of affected (n)	Population	R^2 CAG length only	R^2 CAG + Cand gene	p Value	Protective genotype(s)	Detrimental genotype(s)	Difference in average age of onset (years)	p Value	Ref.
HD	Δ2642	293	England	0.693	0.693	NS					6
		84	France				WT/WT	All others	6.5	<0.05	14
		126	Italy							NS	27
		77	India	0.68	0.68	NS					15
		518	HD MAPS**				WT/Δ2642	Δ2642/Δ2642	0.2	NS	28
		46	Russia		0.764*	0.0067	Δ2642	WT			29
	$(CCG)_n$	77	India	0.68	0.69	0.011					15
		84	France				7/X	7/7	0.3	NS	14
		137	Italy							NS	27
		38	Taiwan	0.50	0.53	NS					30
		54	Russia			NS					29
	Shorter $(CAG)_n$	754	HD MAPS**	0.673	0.680	0.012	Higher repeat lengths	Lower repeat lengths			31
		138	Wales				Lower repeat lengths	Higher repeat lengths		=0.014†	19
		77	India	0.68	0.68	NS					15
		57	Russia		0.677*	NS					29
GRIK2	$(TAA)_n$	293	England	0.693	0.734	<0.008	All others	155, 140			6
		258	US	0.743	0.749	0.009	All others	155	5.5		16
		77	India	0.68	0.70	0.009					15
		276	France	0.684	0.706	0.055					17
		524	Italy	0.68	0.81	0.001‡					32
APOE	ε2/ε3/ε4	293	England	0.693	0.700	NS					6
		60	Greece				ε3/ε4	ε3/ε3	13.8	<0.0002	18
		138	Wales				ε3/ε3 + ε3/ε4	ε2/ε3	11.4	<0.025§	19
		54	Russia		0.695*	0.0542	ε2/ε3	All others			29
		145	Germany	0.272¶		NS	ε4/ε4	All others	1.6	N.S.	33
TCERG1	$(Gln\text{-}Ala)_n$	432	US	0.628	0.632	0.043	304 bp	313 bp			20
		77	India				B (38 repeats?)	C (unknown length)	15.6	=0.0013	15
UCHL1	Ser18Tyr	138	France	0.684	0.726	0.024	Ser/Ser	Ser/Thr + Thr/Thr	1.8	N.S.	17
TP53	Arg72Pro	77	India	0.683	0.723	0.001	Arg/Pro + Pro/Pro	Arg/Arg	1.2	NS	21
DFFB	Arg196Lys	77	India	0.683	0.702	0.007	Arg/Arg	Arg/Lys + Lys/Lys	3.0	NS	21
GRIN2B	C/T SNP	167	Germany	0.308¶	0.346	0.001	C/C	T/T	5.1		22
	T/G SNP	167	Germany	0.308¶	0.324	0.03	T/T	G/G	1.4		22
GRIN2A	C/T SNP	167	Germany	0.308¶	0.322	0.039	C/C	T/T	2.5		22

Multiple regression comparisons are carried out using natural log-transformed age of onset except where noted. Non-significant results (NS) indicate p values >0.1.
*Using linear regression with untransformed age of onset.
†In males only (56 affected). p Value is for linear regression of normal CAG length and age of onset.
‡In juvenile-onset people only (57 affected).
§In males only (56 affected).
¶All cases had between 41 and 45 CAG repeats.
**Drawn from clinics in North America, Europe and Australia.

Table 2 Replication of genetic association tests for age of onset of Huntington's disease in Venezuelan kindreds

Candidate gene	Polymorphism	Number of affected analysed (n)	R^2 CAG length only	R^2 CAG + Cand. gene	p Value for multiple regression	Protective genotype(s)	Detrimental genotype(s)	Difference in average age of onset (years)	p Value for t-test	Phase matches?	p Value for QTDT
HD	Δ2642	375	0.707	0.707	NS	WT {96%}	Δ2642 {4%}	0.04	NS	Yes	NS
	$(CCG)_n$	306	0.747	0.750	NS	$(CCG)_7$ {92%}	$(CCG)_{10}$ {5%}	2.0	NS	Yes	NS
	Shorter $(CAG)_n$	443	0.729	0.730	NS	Lower repeat lengths	Higher repeat lengths		NS	No and Yes*	NS
GRIK2	$(TAA)_n$	368	0.741	0.745	NS	All others	$(TAA)_{16}$	4.6	NS	Yes	0.09 for $(TAA)_{13}$
APOE	ε2/ε3/ε4	423	0.679	0.680	NS	ε3 {77%}	ε4 {10%}	2.0	NS	No	NS
TCERG1	$(Gln\text{-}Ala)_n$	427	0.727	0.733	0.08	306 bp	All others	5.1	NS	Yes	0.07 for 306 bp
UCHL1	Ser18Tyr	405	0.731	0.734	NS	Ser {87%}	Thr {13%}	0.3	NS	Yes	NS
TP53	Arg72Pro	429	0.729	0.729	NS	Arg {62%}	Pro {38%}	0.7	NS	No	NS
DFFB	Arg196Lys	429	0.728	0.730	0.10	Lys {10%}	Arg {90%}	3.0	NS	No	0.02
GRIN2B	C/T SNP	337	0.747	0.747	NS	C {74%}	T {26%}	1.7	NS	Yes	NS
	T/G SNP	305	0.756	0.757	NS	G {49%}	T {51%}	1.1	NS	No	NS
GRIN2A	C/T SNP	421	0.729	0.732	0.04	T {76%}	C {24%}	2.6	NS	Yes	0.07
						T/T	T/C	3.9	0.05	Yes	

*Divergent effects have been reported in different publications (table 1).
All statistical tests controlled for the effect of CAG repeat length on age of onset: the comparisons of R^2 values starts with the correlation of CAG repeat length alone; the comparisons of mean ages of onset for allelic or genotypic classes used CAG repeat length as a covariate, and the QTDT test used residual ages of onset that removed the effect of CAG repeat length. Overall allele frequency in the Venezuelan sample is noted in braces in the "Protective genotype" and "Detrimental genotype" columns, or in tables 3 (for *GRIK2*) and 4 (for *TCERG1* and *GRIN2A*). Non-significant results (NS) indicate p Values >0.1.

considerable effect in the opposite direction, with shorter repeat lengths in the normal *HD* gene being associated with later age of onset ($p = 0.014$).[19] No correlation was seen in either a study of 293 patients in England[6] or a study of 138 patients from Italy[27] (table 1). We also fail to see any effect of the normal *HD* allele in the Venezuelan sample, whether added independently or in interaction with the expanded allele (table 2).

Association with *GRIK2*

Various experimental data suggest that glutamate excitotoxicity may have a role in the pathology of Huntington's disease,[34] which has fuelled interest in glutamate receptor subunits as candidate genes for association with age of onset of Huntington's disease. Association with three glutamate receptor subunits has been reported: *GRIK2*,[6 15 16] *GRIN2B*[22] and *GRIN2A*.[22]

The association of age of onset of Huntington's disease with a $(TAA)_n$ triplet repeat polymorphism in the 3′ untranslated region of *GRIK2*, the GluR6 kinate type glutamate receptor, has the most supportive evidence of all polymorphisms tested (table 1). The study of 293 English patients was the first to document this association, finding that inclusion of the *GRIK2* genotypes in the model for age of onset increased the R^2 statistic from 0.693 to 0.734 ($p<0.008$).[6] A subsequent study in 258 patients from New England confirmed this association, finding that people with the 155 allele of the $(TAA)_n$ polymorphism had an earlier age of onset than those without a 155 allele (32.6 years with compared to 38.1 years without).[16] The R^2 statistic of the exponential regression rose from 0.743 on the basis of the CAG repeat length alone to 0.749 with the $(TAA)_n$ genotype added to the model ($p = 0.009$).[16] Similarly, in the study of 77 Eastern Indian patients, the R^2 statistic of the exponential regression rose from 0.68 to 0.70 with the TAA repeat genotypes included ($p = 0.009$).[15] The same effect falls just shy of significance in a study of 276 patients from France, with the R^2 statistic rising from 0.684 to 0.706 ($p = 0.055$).[17] Finally, although not significant in 524 people from Italy, the R^2 statistic of the exponential regression rose from 0.68 to 0.81 in the 57 juvenile patients in the cohort.[32] This last study reports a significant linear relationship between residual age of onset and the length of the TAA triplet repeat in the *GRIK2* gene, and also shows that most of this effect is driven by people who received a mutant allele from a father, whereas no association is evident in people who received a mutant allele from a mother.[32] This parental effect is not reported in any of the other studies.

Despite the robustness of the effect in other populations, we fail to see strong evidence of association of *GRIK2* in the Venezuelan sample (table 2). Although the R^2 statistic rises from 0.741 to 0.745, the difference does not reach statistical significance. Interestingly, the highest risk allele in Venezuela is the same as the highest risk allele in previous studies. The lowest mean allelic class age of onset in Venezuela is for $(TAA)_{16}$, which corresponds to the high risk 155 allele reported in the previous association studies (table 3). This effect does not reach statistical significance, however, as it is based only on the three people in our affected sample who carry the 16-repeat allele. The population frequency of the 16-repeat allele is eightfold lower in Venezuela than in the other populations investigated (table 3). Hence, the $(TAA)_{16}$ allele would likely show significant evidence of association with reduced age of onset in the Venezuelan population if there were more people with this genotype available for analysis.

Association with *APOE*

The three alleles of *APOE*, apolipoprotein E, are well characterised modifiers of age of onset for Alzheimer's disease.

Table 3 Allele frequencies for the *GRIK2* TAA repeat length polymorphism

Repeat number	Venezuela mean onset	Venezuela frequency (%)	England frequency (%)	US frequency (%)	"Allele number"
9			0.2		134
10	–	0.2			137
11			0.4	0.4	140
12	37.0	4.8	5.2	6.4	143
13	33.0	19.6	8.5	13.6	146
14	34.9	52.9	66.2	63.0	149
15	32.6	22.0	16.7	13.6	152
16	29.7	0.3	2.5	2.7	155
17	41.0	0.2	0.4	0.2	158
18				0.2	161

GRIK2 TAA repeat lengths were determined for 629 people (359 of whom are affected with Huntington's disease) in the Venezuelan kindreds. Allele frequency in all 629 and allelic age of onset for the 359 affected people are given. No affected people had a 10-repeat allele. Allele number but not repeat length number was reported for the samples from England[6] and the US.[16] Assuming the allele frequencies in the three populations are comparable, the repeat numbers and allele numbers correspond as shown in the table, with the major 14-repeat allele in Venezuela being identical to allele 149 in the studies from England and the US. The key risk allele in the previous studies is allele number 155 and it is presumed to be identical to the 16-repeat allele. This key risk allele is 8–9-fold more prevalent in the samples from England and the US than in those from Venezuela.

Carriers with one or two ϵ4 alleles have significantly earlier onset of Alzheimer's disease.[35] A study of 60 patients from Greece tested the association between *APOE* and age of onset for Huntington's disease and found the opposite effect on age of onset for Alzheimer's disease: people with an ϵ3/ϵ4 genotype had later age of onset than those with an ϵ3/ϵ3 genotype ($p<0.002$).[18] The study of 138 patients from Wales also found a similar effect of *APOE* genotype on age of onset of Huntington's disease, finding that an ϵ2/ϵ3 genotype is associated with earlier age of onset, though in males only ($p<0.025$).[19] These studies are in contrast with those on 293 patients in England[6] and 145 patients in Germany,[33] which examined *APOE* genotypes but found no considerable effect on age of onset (table 1).

This association is not present in the Venezuelan sample either (table 2). People in the Venezuelan cohort with an ϵ4/ϵ4 genotype have a non-significant earlier age of onset for Huntington's disease, an effect contrary to those of the previous two studies. To mimic the analysis of sex-specific associations in the Welsh study, we further analysed each combination of allele or genotype with sex. The only trend towards significance we find is that women with an ϵ3 allele have a slightly later age of onset compared with everyone else ($p = 0.04$). We do not consider this replication of association as it does not match the previous result of the ϵ2/ϵ3 genotype being associated with earlier age of onset in men only.

Table 4 Allele frequencies for polymorphisms in *TCERG1* and *GRIN2A*

Gene	Allele number or genotype	Presumed $(Gln\text{-}Ala)_n$ repeat number	Population allele frequency (%)	Mean age of onset
TCERG1	289	35	2.5	29.3
	294	36	0.8	41.4
	300	37	0.2	38.5
	306	38	95.4	34.8
	318	40	0.7	26.3
	328	41.5	0.3	–
GRIN2A	C/C		4.7	33.8
	C/T		39.5	32.3
	T/T		55.8	36.2

TCERG1 genotypes were determined for 728 people (410 of whom are affected with Huntington's disease). No affected people had a *TCERG1* 328 allele. Repeat length was inferred on the basis of the length of the amplified PCR products and the published sequence. *GRIN2A* genotypes were determined for 731 people (405 of whom were affected with Huntington's disease). Allele frequencies are calculated from all people. Mean age of onset is calculated from affected people.

Association with *TCERG1*

TCERG1, the transcriptional co-activator CA150, is the human homologue of a *Caenorhabditis elegans* protein that interacts with N-terminal fragments of huntingtin.[20] Analysis of 432 American patients with Huntington's disease showed a significant association of an imperfect $(Gln\text{-}Ala)_n$ repeat polymorphism in the protein ($p = 0.043$).[20] The study of 77 patients from India also reported a significant association ($p = 0.013$)[15] (table 1).

We find a trend towards evidence for association at this gene. A categorical predictor coding for the genotypic classes of this gene causes the R^2 statistic to rise from 0.727 to 0.733 ($p = 0.08$). Moreover, the orthogonal model of association implemented in QTDT also provides some evidence for association (table 2) with the 306-bp allele of *TCERG1* ($p = 0.07$), suggesting that this allele is transmitted with skewed frequency in the 64 members informative for this analysis. The low significance of this potential association may be due in part to the uninformative nature of the polymorphism, with a 95.4% allele frequency of the major 306-bp allele (table 3).

Association with *UCHL1*

A rare mutation in *UCHL1* causes a familial form of Parkinson's disease.[36] The study with 276 French patients showed association with age of onset of Huntington's disease for one of the two more common polymorphisms in this gene that they tested, a serine to threonine missense mutation ($p = 0.024$)[17] (table 1). In the Venezuelan sample, the R^2 statistic increases from 0.731 to 0.734 with the *UCHL1* genotypes included, but this effect is not statistically significant (table 2).

Association with *TP53* and *DFFB*

The study of 77 patients from India showed significant association with two candidate polymorphisms, an arginine to proline polymorphism in *TP53* ($p = 0.001$) and an arginine to lysine polymorphism *DFFB* ($p = 0.007$)[21] (table 1).

In the Venezuelan sample, we found no evidence for association with the TP53 polymorphism. We see a trend towards a significant increase in fitness when DFFB genotypes are included, with the R^2 statistic increasing from 0.728 to 0.730 ($p = 0.10$). We also get a significant result from association analysis using QTDT ($p = 0.02$). Despite these results, we do not consider it a true replication of the original

study because the protective and at-risk alleles are opposite in the two studies: the lysine allele is protective in the Venezuelan sample (table 2), whereas the arginine allele is protective in the Indian sample (table 1).

Association with *GRIN2B* and *GRIN2A*

A study with 308 patients from Germany showed association with three SNPs in two genes encoding glutamate channel subunits: a C/T SNP ($p = 0.03$) and a T/G SNP ($p = 0.03$) in *GRIN2B*, and a C/T SNP ($p = 0.039$) in *GRIN2A*[22] (table 1).

No evidence was obtained for association at either of the *GRIN2B* SNPs in the Venezuelan sample. We did, however, find evidence of association of the *GRIN2A* SNP (table 2). The R^2 statistic rose modestly (from 0.729 to 0.732) but significantly ($p = 0.04$) when *GRIN2A* genotypes were added to the regression model. Furthermore, although the 2.6-year difference in the mean age of onset for the C and T allelic classes did not reach statistical significance, the 3.9-year difference in the mean age of onset between the C/C and T/C genotype classes was statistically significant (tables 2 and 4). This result is further supported by a trend ($p = 0.07$) for the QTDT association test (table 2).

The *GRIN2A* C/T SNP is of some general interest as heterozygosity varies greatly by geographical region, with the C allele frequency at 17% in Europe and at 52–56% in Asia and Africa.[37] The Venezuelan C allele frequency is 24% (table 4).

DISCUSSION

Our goal is to obtain a complete understanding of the 17% variance in age of onset that is attributable to genetic factors other than the length of the CAG repeat in the *HD* gene.[12] Our first step towards this goal was to determine whether other genes previously reported to affect the age of onset for Huntington's disease also have an effect in the Venezuelan kindreds. We tested 12 polymorphisms that gave a statistically significant association (ie, with $p<0.05$) with age of onset of Huntington's disease in at least one other population.

Although sensitive statistical tests measure the genetic association using case–control comparisons (such as using changes in R^2 to assess the amount of total variance in age of onset explained by a given polymorphism or comparing the mean age of onset between genotypic or allelic class groups), they are not ideally suited to groups such as the Venezuelan kindreds, where the people are not entirely genetically independent of one another. Still, we believe that the Venezuelan kindreds are large enough to make the deviation due to inter-relatedness negligible.

Of the 12 polymorphisms we tested, only *GRIN2A* gave consistent evidence of replication in the Venezuelan sample at the same level of significance. *GRIN2A* encodes the NR2A subunit of the NMDA-type glutamate receptor. This subunit is expressed on the striatal medium spiny neurone, the cell type that is most prone to degeneration in Huntington's disease.[38] These cells also seem to be particularly sensitive to glutamate excitotoxicity, as injection of quinolinic acid (an excitotoxic agent) into the striatum of rodents or primates causes the same neurones to die, resulting in a Huntington's disease-like syndrome.[39] Variation in function or expression of glutamate receptor subunits could modulate excitotoxic death and affect the age of onset.

Although it does not achieve statistical significance (with $p<0.05$), a repeat length polymorphism in the imperfect Gln–Ala repeat of *TCERG1* shows a trend towards association with age of onset of Huntington's disease in the Venezuelan sample ($p<0.10$). Both the multiple regression model and the QTDT orthogonal test show a trend towards association (table 2). The *TCERG1* protein interacts with huntingtin,[20] and the strength of this interaction could certainly be modulated by the repeat length polymorphism in *TCERG1*.

Although each explains only a small portion of the total variance in age of onset, statistically significant or near-significant associations between age of onset of Huntington's disease and polymorphisms in *GRIN2A* and *TCERG1* have been shown in all studies that reported analysing these genes. Two of two published studies report an association for *GRIN2A,* (Arning et al[22] and this study) and three of three published studies and this study report at least near-significant association for *TCERG1* (Chattopadhyay et al,[15] Holbert et al,[20] and this study). Although the lack of negative reports may be in part owing to publication bias for positive association results, both *GRIN2A* and *TCERG1* deserve additional attention by those interested in the genetics related to Huntington's disease.

We failed to replicate genetic association with the remaining 10 polymorphisms (table 2). Nine of the statistical tests failed replication because they produced p values >0.1 (table 2). The remaining association test (for *DFFB*) showed a trend towards significance, but was still considered a failed replication test because the direction of the effect was contrary to those reported in the initial study.

The lack of replication in the Venezuelan sample for these 10 association tests does not imply that the original association results are not genuine. The complexity of interpreting genetic association tests is well known.[40] A true association will not always replicate in a new population, as the allele frequencies of the polymorphisms at the key gene must also be conducive to detecting the association. For example, although five independent populations with Huntington's disease show significant or near-significant association with the TAA repeat in the *GRIK2* gene,[6 15–17 32] we see no significant evidence for association in Venezuela (table 2). In the previous studies from England[6] and the US,[16] the $(TAA)_n$ genotype conferring the greatest risk is the 16-repeat allele. We see the same effect in our sample as well: the 16-repeat allele confers the greatest risk for early onset (table 3). Because the allele frequency is so low in the Venezuelan sample, however, this effect is not statistically significant.

This study moves us closer towards our goal of a complete understanding of the variability in age of onset of Huntington's disease. We hope that a full accounting of all the genes that regulate the age of onset of Huntington's disease will advance our understanding of the aetiology of the disease, while at the same time providing useful therapeutic targets for drug discovery.

ACKNOWLEDGEMENTS

We thank families with Huntington's disease in Venezuela and throughout the world. We also acknowledge grants from the NINDS, NIH, the WM Keck Foundation, and the Hereditary Disease Foundation.

Authors' affiliations

J M Andresen, G Alkorta-Aranburu, E A Addis, D E Housman, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
J Gayán, D Brocklebank, L R Cardon, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
S S Cherny, Department of Psychiatry and Genome Research Centre, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong
N S Wexler, Columbia University, New York, New York, USA

Funding: The study was supported by grants from the NINDS, NIH, WM Kneck Foundation and Hereditary Foundation.

Competing interests: None.

REFERENCES

1 **Bates G**, Harper P, Jones L. *Huntington's disease, (*3rd edn., Oxford, UK: Oxford University Press 2002.

2 **Huntington's Disease Collaborative Research Group**. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993;**72**:971–83.
3 **Brinkman RR**, Mezei MM, Theilmann J, *et al*. The likelihood of being affected with Huntington disease by a particular age, for a specific CAG size. *Am J Hum Genet* 1997;**60**:1202–10.
4 **Duyao M**, Ambrose C, Myers R, *et al*. Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nat Genet* 1993;**4**:387–92.
5 **Ranen NG**, Stine OC, Abbott MH, *et al*. Anticipation and instability of IT-15 $(CAG)_n$ repeats in parent-offspring pairs with Huntington disease. *Am J Hum Genet* 1995;**57**:593–602.
6 **Rubinsztein DC**, Leggo J, Chiano M, *et al*. Genotypes at the GluR6 kainate receptor locus are associated with variation in the age of onset of Huntington disease. *Proc Natl Acad Sci USA* 1997;**94**:3872–6.
7 **Snell RG**, MacMillan JC, Cheadle JP, *et al*. Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. *Nat Genet* 1993;**4**:393–7.
8 **Andrew SE**, Goldberg YP, Kremer B, *et al*. The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nat Genet* 1993;**4**:398–403.
9 **Langbehn DR**, Brinkman RR, Falush D, *et al*. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. *Clin Genet* 2004;**65**:267–77.
10 **Rosenblatt A**, Brinkman RR, Liang KY, *et al*. Familial influence on age of onset among siblings with Huntington disease. *Am J Med Genet* 2001;**105**:399–403.
11 **Li JL**, Hayden MR, Almqvist EW, *et al*. A genome scan for modifiers of age at onset in Huntington disease: The HD MAPS Study. *Am J Hum Genet* 2003;**73**:682–7.
12 **Wexler NS**, Lorimer J, Porter J, *et al*. Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proc Natl Acad Sci USA* 2004;**101**:3498–503.
13 **Farrer LA**, Cupples LA, Wiater P, *et al*. The normal Huntington disease (*HD*) allele, or a closely linked gene, influences age at onset of HD. *Am J Hum Genet* 1993;**53**:125–30.
14 **Vuillaume I**, Vermersch P, Destee A, *et al*. Genetic polymorphisms adjacent to the CAG repeat influence clinical features at onset in Huntington's disease. *J Neurol Neurosurg Psychiatry* 1998;**64**:758–62.
15 **Chattopadhyay B**, Ghosh S, Gangopadhyay PK, *et al*. Modulation of age at onset in Huntington's disease and spinocerebellar ataxia type 2 patients originated from eastern India. *Neurosci Lett* 2003;**345**:93–6.
16 **MacDonald ME**, Vonsattel JP, Shrinidhi J, *et al*. Evidence for the GluR6 gene associated with younger onset age of Huntington's disease. *Neurology* 1999;**53**:1330–2.
17 **Naze P**, Vuillaume I, Destee A, *et al*. Mutation analysis and association studies of the ubiquitin carboxy-terminal hydrolase L1 gene in Huntington's disease. *Neurosci Lett* 2002;**328**:1–4.
18 **Panas M**, Avramopoulos D, Karadima G, *et al*. Apolipoprotein E and presenilin-1 genotypes in Huntington's disease. *J Neurol* 1999;**246**:574–7.
19 **Kehoe P**, Krawczak M, Harper PS, *et al*. Age of onset in Huntington disease: sex specific influence of apolipoprotein E genotype and normal CAG repeat length. *J Med Genet* 1999;**36**:108–11.
20 **Holbert S**, Denghien I, Kiechle T, *et al*. The Gln-Ala repeat transcriptional activator CA150 interacts with huntingtin: neuropathologic and genetic evidence for a role in Huntington's disease pathogenesis. *Proc Natl Acad Sci USA* 2001;**98**:1811–16.
21 **Chattopadhyay B**, Baksi K, Mukhopadhyay S, *et al*. Modulation of age at onset of Huntington disease patients by variations in TP53 and human caspase activated DNase (hCAD) genes. *Neurosci Lett* 2005;**374**:81–6.
22 **Arning L**, Kraus PH, Valentin S, *et al*. NR2A and NR2B receptor gene variations modify age at onset in Huntington disease. *Neurogenetics* 2005;**6**:25–8.
23 **Paschen W**, Blackstone CD, Huganir RL, *et al*. Human GluR6 kainate receptor (GRIK2): molecular cloning, expression, polymorphism, and chromosomal assignment. *Genomics* 1994;**20**:435–40.
24 **Pramanik S**, Basu P, Gangopadhaya PK, *et al*. Analysis of CAG and CCG repeats in Huntingtin gene among HD patients and normal populations of India. *Eur J Hum Genet* 2000;**8**:678–82.
25 **Wigginton JE**, Abecasis GR. PEDSTATS: descriptive statistics, graphics, and quality assessment for gene mapping data. *Bioinformatics* 2005;**21**:3445–7.
26 **Abecasis GR**, Cookson WO, Cardon LR. Pedigree tests of transmission disequilibrium. *Eur J Hum Genet* 2000;**8**:545–51.
27 **Novelletto A**, Persichetti F, Sabbadini G, *et al*. Polymorphism analysis of the huntingtin gene in Italian families affected with Huntington disease. *Hum Mol Genet* 1994;**3**:1129–32.
28 **Djousse L**, Knowlton B, Hayden MR, *et al*. Evidence for a modifier of onset age in Huntington disease linked to the HD gene in 4p16. *Neurogenetics* 2004;**5**:109–14.
29 **Kutuev I**, Khusainova R, Khidiyatova I, *et al*. Analysis of the IT15 gene in Huntington's disease families. *Genetika* 2004;**40**:1123–30.
30 **Wang CK**, Wu YR, Hwu WL, *et al*. DNA haplotype analysis of CAG repeat in Taiwanese Huntington's disease patients. *Eur Neurol* 2004;**52**:96–100.
31 **Djousse L**, Knowlton B, Hayden M, *et al*. Interaction of normal and expanded CAG repeat sizes influences age at onset of Huntington disease. *Am J Med Genet* 2003;**119A**:279–82.
32 **Cannella M**, Gellera C, Maglione V, *et al*. The gender effect in juvenile Huntington disease patients of Italian origin. *Am J Med Genet B Neuropsychiatr Genet* 2004;**125**:92–8.
33 **Saft C**, Andrich JE, Brune N, *et al*. Apolipoprotein E genotypes do not influence the age of onset in Huntington's disease. *J Neurol Neurosurg Psychiatry* 2004;**75**:1692–6.
34 **Feigin A**, Zgaljardic D. Recent advances in Huntington's disease: implications for experimental therapeutics. *Curr Opin Neurol* 2002;**15**:483–9.
35 **Hardy J**. Apolipoprotein E in the genetics and epidemiology of Alzheimer's disease. *Am J Med Genet* 1995;**60**:456–60.
36 **Leroy E**, Boyer R, Auburger G, *et al*. The ubiquitin pathway in Parkinson's disease. *Nature* 1998;**395**:451–2.
37 **Database of Single Nucleotide Polymorphisms (dbSNP)**. Bethesda, MD: National Center for Biotechnology Information, National Library of Medicine. dbSNP accession: {rs1969060}, {dbSNP Build ID, {36.1}}.http://www.ncbi.nlm.nih.gov/SNP/ (accessed 7 Oct 2006).
38 **Young AB**, Greenamyre JT, Hollingsworth Z, *et al*. NMDA receptor losses in putamen from patients with Huntington's disease. *Science* 1988;**241**:981–3.
39 **Ellison DW**, Beal MF, Mazurek MF, *et al*. Amino acid neurotransmitter abnormalities in Huntington's disease and the quinolinic acid animal model of Huntington's disease. *Brain* 1987;**110**:1657–73.
40 **Cardon LR**, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001;**2**:91–9.