Table e-1: Baseline characteristics

Total no. of participants	3411
Sex [% female]	50.2%
Age of onset [yrs]	48.7 (11.3)
Disease duration [yrs]	2.8 (3.6)
UHDRS TFC score	8.0 (3.3)
UHDRS TMS score	37.6 (18.8)
BMI	25.2 (4.9)
Mutant CAG repeat size	43.9 (3.1)

Data are represented as mean ± standard deviation.

Abbreviations: BMI = body mass index, TFC = Total Functional Capacity, TMS = Total Motor Score, UHDRS = Unified Huntington Disease Rating Scale.

Table e-2: Simulation and analysis strategy

	Description	Model
Step 1:	Estimation of the effect of CAG repeat size on age of onset (AO_i) for subject i using the following linear regression equation:	$ln(AO_i) = \beta_0 + \beta_1 \cdot CAG_i + \varepsilon_i$
Step 2:	Residual age of onset (RAO_i), not explained by CAG repeat size, was defined as the difference between the actual age of onset (AO_i) and predicted age of onset ($\widehat{AO}_i$) using the formula described in step 1, after back transformation into the natural scale.	$RAO_i = AO_i - \widehat{AO}_i = AO_i \cdot (1 - e^{-\varepsilon_i})$
Step 3:	For each clinical score, including total functional capacity (TFC), total motor score (TMS) and a cognitive summary score (PC1), we first estimated the average effect of age (as a measure of disease duration) on the rate of progression using the actual Enroll-HD dataset. In this model, $Y_i(t)$ represents the clinical score at time t for subject i, t denotes age in years, β_0 represents the average intercept, β_1 denotes the mean rate of disease progression averaged over the entire cohort, $b_{0,i}$indicates the random intercept for subject i, while $b_{1,i}$ denotes the random slope for subject i (which can be interpreted as the difference between the rate of disease progression for subject i and the average rate of disease progression) while ξ_i represents a random residual error term. We used REML estimation with an unstructured covariance matrix.	$Y_i(t) = \beta_0 + \beta_1 \cdot t + b_{0,i} + b_{1,i} \cdot t + \xi_i$
Step 4:	Simulation of a scenario in which the rate of disease progression in each mutation carrier is determined by his or her CAG repeat size (CAG_i), residual age of onset (RAO_i), some other (unknown) factors (U_i) and a random error term (ξ_i). In this model, $\tilde{C}_i$, $\tilde{R}_i$and $\tilde{U}_i$denote standardized vectors (i.e. with a mean of zero and a standard deviation of one) of CAG_i, RAO_i and U_i, respectively. The fraction of the variation in disease progression determined by the combination of $CAG_i$and RAO_i is modelled by the f_1 parameter, while the fraction of $f_1$which is determined by CAG_i alone is denoted by the f_2 parameter. For the simulation experiments we based estimates of β_0 and β_1 as well as estimates of the variation in the random slopes (i.e. $Var(b_0)$) and errors (i.e. $Var(\xi)$) on the estimates obtained in step 3 from the actual patient data (note that, by definition, both b_0 and ξ have a mean of zero). For each clinical score (including TFC, TMS and PC1) we simulated four different scenarios in which mutant *HTT* CAG repeat size and residual age of onset were precisely modelled to explain 0, 25%, 50%, 75% or 100%, respectively, of the variation in the rate of disease progression (i.e. $f_1 = \{0, 0.25, 0.5, 0.75, 1\}$). For simplicity, we assumed that in each scenario CAG repeat size and residual age of onset both contributed equally to the variation in the rate of disease progression, i.e. $f_2 = 1 - f_2 = 0.5$.	$Y_i(t) = \beta_0 + \beta_1 \cdot (\sqrt{f_1}(\sqrt{f_2}\tilde{C}_i + \sqrt{1 - f_2}\tilde{R}_i) + \sqrt{1 - f_1}\tilde{U}_i + 1) \cdot t + b_{0,i} + \xi_i$
Step 5:	To validate that the variation in disease progression in the simulated scenarios was indeed modelled correctly according to the prespecified parameters, we analyzed the simulated datasets with the mixed-effects models described under step 3 and determined the variation in the rate of disease progression accounted for by CAG repeat size (R^2_{CAG}) and residual age of onset (R^2_{RAO}) by calculating the proportion of decrease in variance of the random slope term by sequentially adding $\tilde{C}_i$ and its interaction with t, followed by $\tilde{R}_i$ and its interaction with t (specified models) as compared to a model with only age as a predictor (null model).	$R^2 = 1 - \frac{Var(b_1)\ in\ specified\ model}{Var(b_1)\ in\ null\ model}$

Table e-3: Simulation results

	f_1*	R^2_{CAG}†	R^2_{RAO}†	$R^2_{CAG+RAO}$†	Difference‡
Total functional capacity	0	0.00	0.00	0.00	0.00
	0.25	0.16	0.13	0.30	0.05
	0.50	0.29	0.24	0.55	0.05
	0.75	0.40	0.35	0.78	0.03
	1	0.52	0.46	1.00	0.04
Total motor score	0	0.00	0.00	0.00	0.00
	0.25	0.14	0.14	0.28	0.03
	0.50	0.26	0.25	0.53	0.03
	0.75	0.37	0.37	0.77	0.02
	1	0.49	0.48	1.00	0.02
Cognitive summary score	0	0.00	0.00	0.00	0.00
	0.25	0.16	0.13	0.29	0.04
	0.50	0.28	0.24	0.53	0.03
	0.75	0.40	0.35	0.77	0.02
	1	0.52	0.48	1.00	0.02
Body mass index	0	0.00	0.00	0.00	0.00
	0.25	0.13	0.14	0.25	0.02
	0.50	0.26	0.25	0.48	0.02
	0.75	0.38	0.36	0.69	0.06
	1	0.50	0.45	0.89	0.11

Legend:

*****) The fraction of the variation in disease progression determined by the combination of $CAG_i$and RAO_i is modelled by the f_1 parameter, ranging from 0 to 1. See **Table E-2** for additional details.

†) Note that due to a weak correlation between CAG repeat size and residual age of onset (Pearson's r = -0.03, p < 0.001) the combined coefficient of determination ($R^2_{CAG+RAO}$) is generally slightly different than the sum of the unique variable specific coefficients of determination (R^2_{CAG} and R^2_{RAO}) as it also includes the effect of the covariance between CAG repeat size and RAO.

‡) This column contains the maximal difference between the prespecified coefficients of determination (f_1 and f_2) and the retrieved coefficients of determination. Note that for simplicity f_2 , the proportion of variation modelled to be due to CAG repeat size alone, was set to 0.5.

Table e-4: Sensitivity analysis: The association between *HTT* CAG repeat size, residual age of onset and clinical progression in HD in the total cohort without excluding outliers and irrespective of BMI.

	Age[1]	CAG[2]	CAG × age[3]	RAO[4]	RAO × age[5]	R^2_{CAG}[6]	R^2_{RAO}[6]	$R^2_{CAG+RAO}$[6]
Total functional capacity	-5.56×10^{-1} (-5.70×10^{-1} to -5.42×10^{-1})***	-1.77 (-1.82 to -1.72)***	-2.54×10^{-2} (-2.78×10^{-2} to -2.30×10^{-2})***	3.63×10^{-1} (3.46×10^{-2} to 3.81×10^{-2})***	-3.95×10^{-3} (-5.03×10^{-3} to -2.87×10^{-3})***	40.4 (36.9 to 44.1)	7.9 (6.2 to 9.4)	62.6 (58.8 to 66.2)
Total motor score	3.57 (3.49 to 3.65)***	12.16 (11.85 to 12.47)***	1.89×10^{-1} (1.75×10^{-1} to 2.03×10^{-1})***	-2.04 (-2.14 to -1.94)***	2.10×10^{-2} (1.47×10^{-2} to 2.72×10^{-2})***	46.7 (43.7 to 50.1)	8.0 (6.8 to 9.3)	65.9 (63.0 to 69.3)
Cognitive summary score	-2.49×10^{-1} (-2.57×10^{-1} to -2.41×10^{-1})***	-8.74×10^{-1} ($-9.07\ 10^{-1}$ to -8.40×10^{-1})***	-1.52×10^{-2} (-1.68×10^{-2} to -1.36×10^{-2})***	1.07×10^{-1} (9.57×10^{-2} to 1.17×10^{-1})***	-9.35×10^{-4} (-1.62×10^{-3} to -2.43×10^{-4})***	42.0 (38.5 to 46.1)	2.4 (1.4 to 3.6)	49.7 (45.8 to 53.9)
BMI	-1.04×10^{-1} (-1.24×10^{-1} to -8.32×10^{-2})***	-6.02×10^{-1} (-6.82×10^{-1} to -5.22×10^{-1})***	-1.24×10^{-2} (-1.63×10^{-2} to -7.52×10^{-3})***	1.74×10^{-2} (-9.75×10^{-2} to 4.45×10^{-2})	-4.05×10^{-3} (-4.72×10^{-3} to -1.37×10^{-3})***	3.4 (1.6 to 5.2)	0.0 (-0.5 to 0.4)	4.1 (2.2 to 6.0)

Legend: Values represent parametric means and 95% confidence intervals of the mean, except for R^2 (last three column). As the underlying distribution of the R^2 statistic was unknown, for this statistic we calculated bootstrapped means and 95% bias-corrected and accelerated confidence intervals based on 1000 random resamplings with replacement of the original dataset, while taking into account the clustering of the measurements per subject. $^{**}p < 0.01$, $^{***}p < 0.01$.

[1]) This column contains the regression coefficients associated with age which can be interpreted as the rate of disease progression per year in units of the outcome measure.

[2]) This column contains the regression coefficients associated with expanded *HTT* CAG repeat size which can be interpreted as the average increase or decrease in the outcome measure during the follow-up period per one repeat increase.

[3]) This column contains the regression coefficients of the interaction term between expanded *HTT* CAG repeat size and age: A significant interaction means that CAG repeat size affects the rate of disease progression.

[4]) This column contains the regression coefficients associated with residual age of onset (RAO) which can be interpreted as the average increase or decrease in the outcome measure during the follow-up period per one year onset later than expected.

[5]) This column contains the regression coefficients of the interaction term between residual age of onset (RAO) and age: A significant interaction means that RAO affects the rate of disease progression.

[6]) These columns represent the coefficients of determination (in percentages) associated with expanded *HTT* CAG repeat size ($\boldsymbol{R^2_{CAG}}$), residual age of onset ($\boldsymbol{R^2_{RAO}}$) or both ($\boldsymbol{R^2_{CAG+RAO}}$) and can be interpreted as the fraction of variation in disease progression which can be attributed to *HTT* CAG repeat size, residual age of onset or both acting together, respectively. Note that $\boldsymbol{R^2_{CAG+RAO}}$ is higher than the sum of $\boldsymbol{R^2_{CAG}}$ and $\boldsymbol{R^2_{RAO}}$ as these latter two represent estimates of the unique contribution of either *HTT* CAG repeat size or residual age of onset to disease progression, respectively, while the former also includes the proportion of variance explained by their covariance.