Figure e-1. Data quality assessment. Plotting of the standard deviation of age of onset against expanded *HTT* CAG repeat size demonstrated substantial heteroscedasticity **(A)**, which was alleviated for the 40-57 CAG repeat range after a natural logarithmic transformation of age of onset **(B)**. To identify outliers, we plotted the natural logarithm of age of onset against expanded *HTT* CAG repeat size and defined outliers as those points falling either above or below 1.5 times the interquartile range (IQR) of the associated boxplot (small circles represent outliers, thick black horizontal lines represent medians, boxes indicate IQR, while the whiskers denote 1.5 x IQR per CAG repeat category) **(C).**

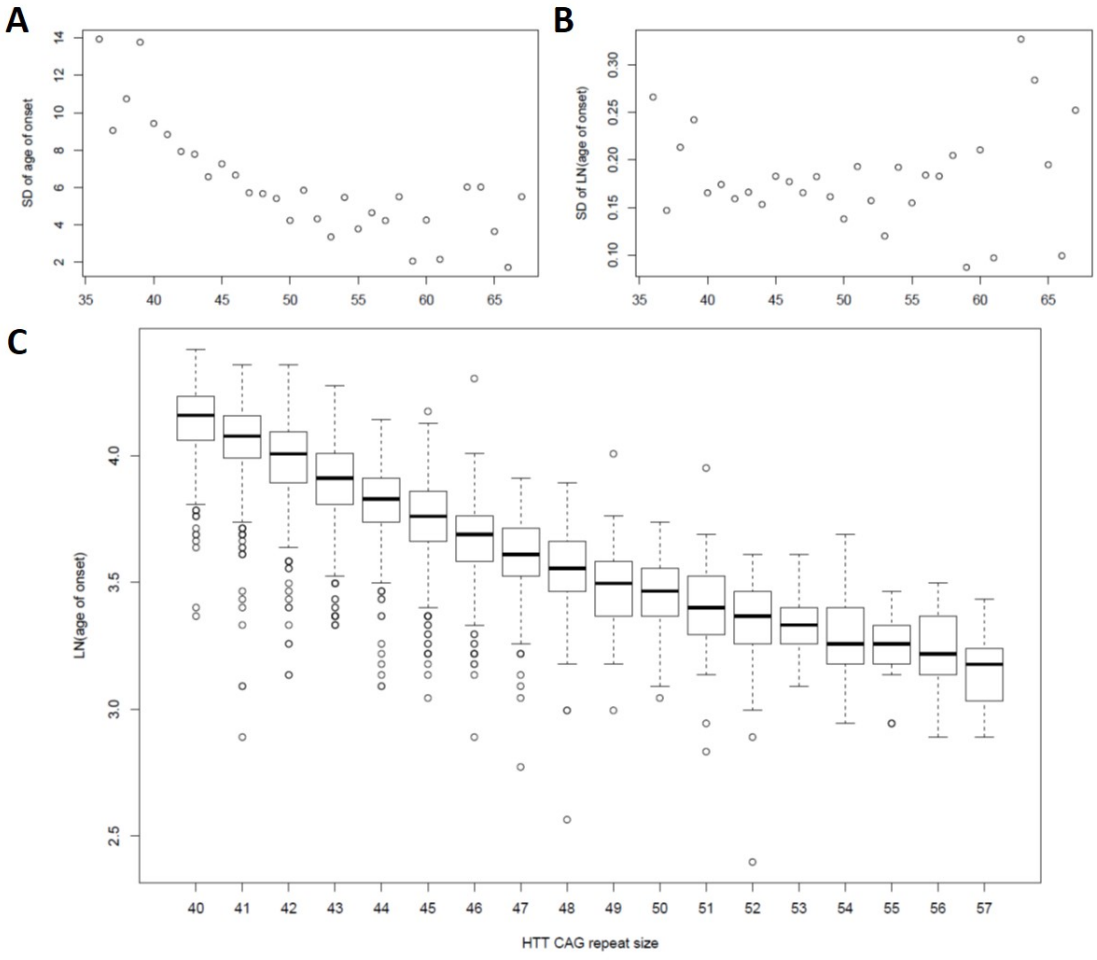

Figure e-2. The relation between age of onset and mutant *HTT* CAG repeat size. We plotted the natural logarithm of age of onset against the number of CAG repeats in the expanded *HTT* allele. The red line represents the regression line, while the shaded areas around the line indicate the 95% confidence interval of the mean. The points represent individual patient data with horizontally added random jitter to better illustrate the density of the data points per CAG repeat category. The regression model explained 69.3% of the total variance in age of onset.

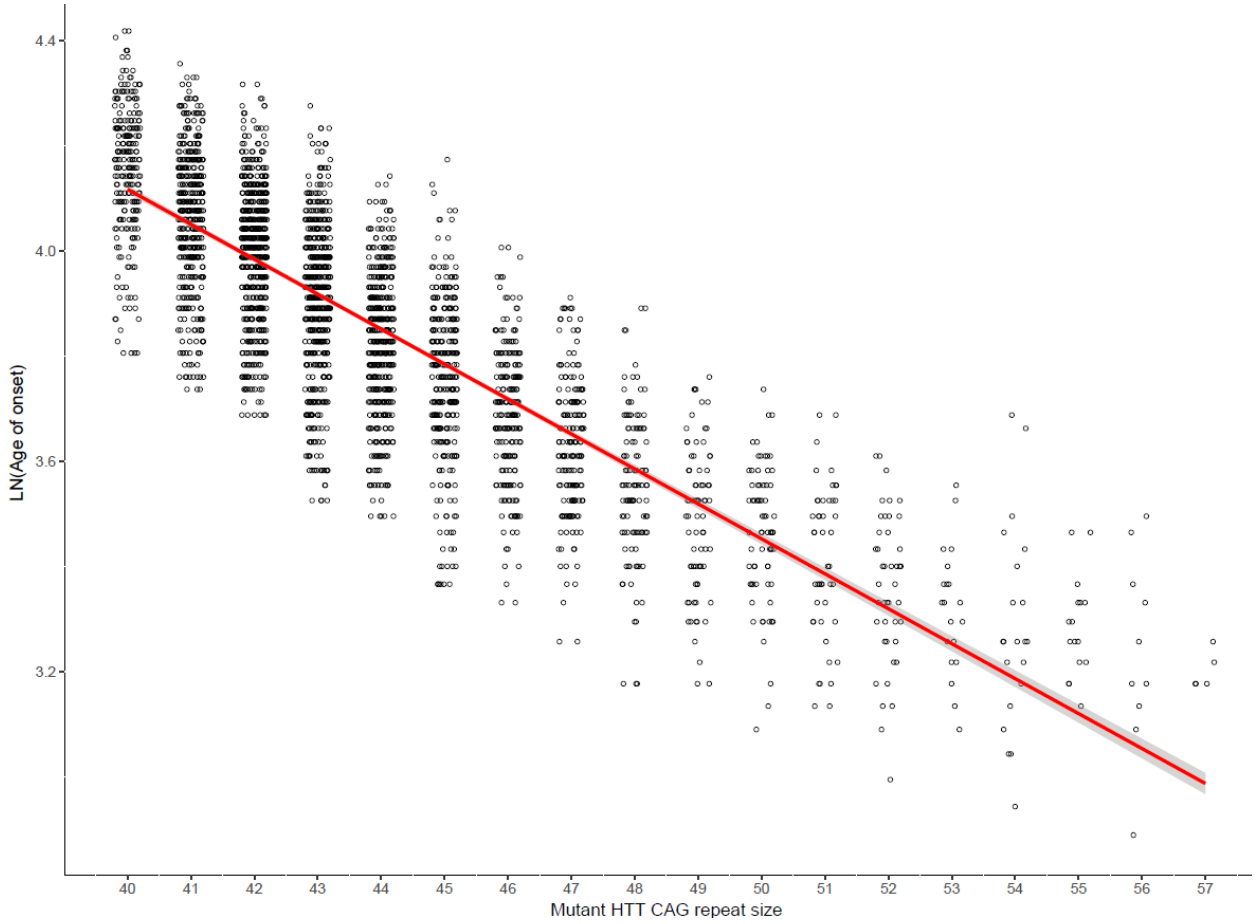

Figure e-3. Regression diagnostics. The regression diagnostics did not show any particular pattern in the distribution of the residuals when plotted against the model predicted values for age of onset based on *HTT* CAG repeat size **(A, B)**, and also exhibited no major departures from normality **(C)** (the largest studentized residual had a t-value of 3.67 with a Bonferroni adjusted p-value of 0.83). Similarly, there were no influential points with the largest Cook's distance being 0.025 **(D).**

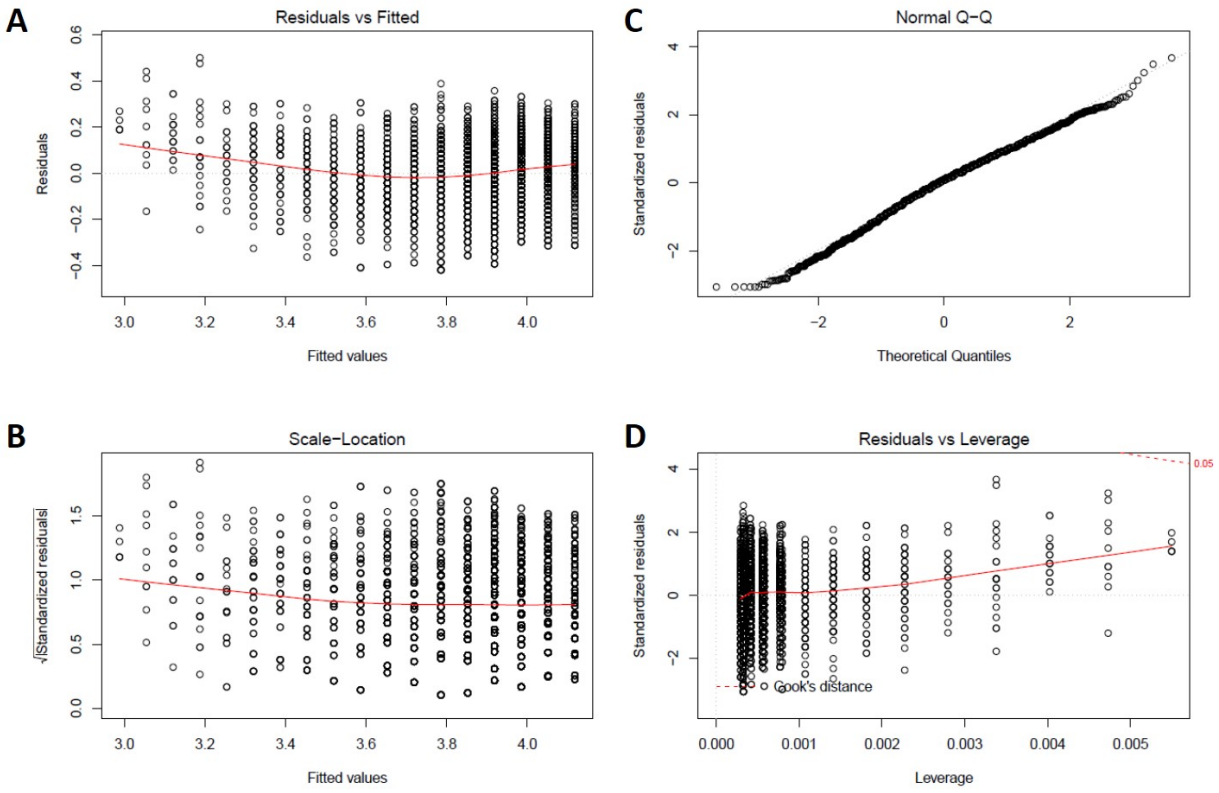

Figure e-4. Distribution of residuals of age of onset on the natural scale. The residuals of age of onset, after regression on expanded *HTT* CAG repeat size, followed a normal distribution on the natural scale (represented by the red curve for reference).

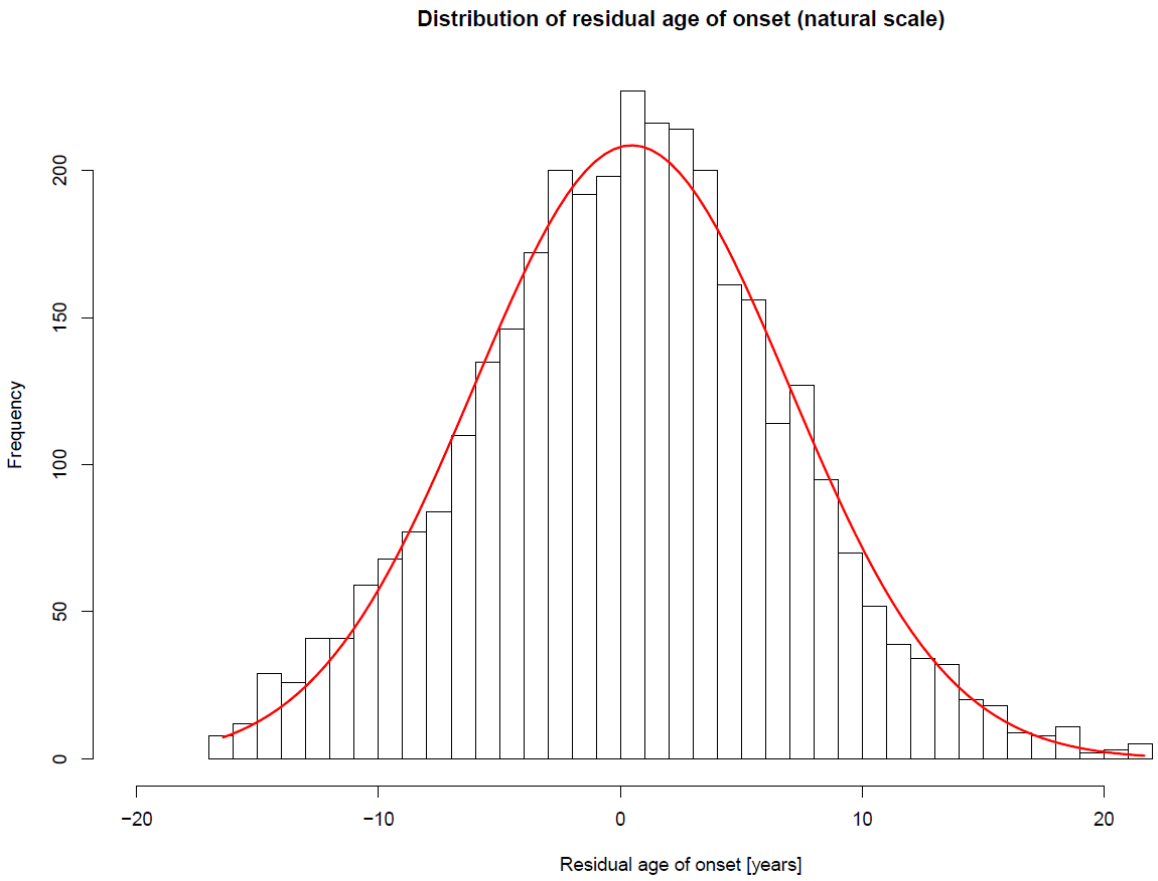

Figure e-5. Distribution of disease duration and age. For every measurement occasion, disease duration was defined as the difference between the participant's age at that occasion and his/her age at onset. Median disease duration, i.e. time from disease onset, was 4.0 years (interquartile range: 2.0 to 7.0 years), while median follow-up time was 2.2 years (interquartile range: 1.2 to 4.3 years). Note the highly right-skewed distribution of disease duration **(A).** On the other hand, participants' age was normally distributed **(B).**

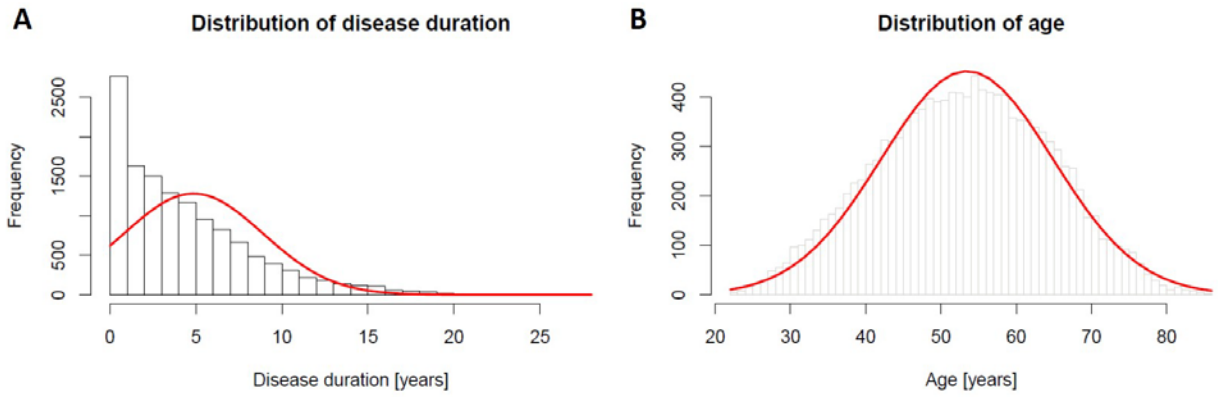

Figure e-6. Model fits. To assess model fit we plotted the measured values for total functional capacity (**A**), total motor score (**B**), cognitive summary score (**C**) and body mass index (BMI) against the predicted values obtained through the linear mixed-effects models. Visual inspection of these plots confirmed a good model fit without major indications for non-linearity, heteroscedasticity or the existence of influential points.