Participants

All participating countries for each site (Paris, Vancouver, Leiden and London) obtained local ethics approval for this study and all participants provided written informed consent. Sites each recruited about 30 controls, 30 premanifest HD (PreHD) and 30 early HD participants. 349 participants participated in this study, including controls (n=116), premanifest (n=117) and early HD (n=116); participants with only baseline data being excluded. Of the 17 participants excluded, 16 withdrew prior to the 12-month visit and 1 control had been diagnosed with Parkinson's disease after the baseline visit. Premanifest and early HD participants were required to have had prior voluntary independent genetic testing for the HD-CAG repeat expansion. Repeat CAG testing was performed for the study using a polymerase-chain-reaction method.[1] The PreHD and early HD participants had ≥36 and controls had <36 CAG repeats. At the 24 month follow up, of 111 premanifest participants assessed at 24 months, 78 remained premanifest, 21 were peri-manifest (i.e., TMS > 5), and 12 were diagnosed with early HD.

Procedures and Cognitive Assessment

Cognitive examiners, typically only one individual per site across all three visits, were trained and certified in standardised procedures by the first author. Quality control procedures included onsite and online data monitoring and automatic range checks within the database. As detailed previously,[2] participants also underwent MRI, neuropsychiatric, oculomotor, quantitative motor and quality of life assessments.

The test selection was based on findings from the Predict-HD study,[3-5] which included a large cognitive battery and focused on PreHD, and on findings from an extensive systematic review and meta-analysis of the literature on cognition in PreHD and early HD conducted by our research group. Participants were tested in the language spoken locally at each site. To minimise the effects of using samples being tested in different languages, we selected tests that did not have a large verbal component and all instructions and verbal test stimuli were checked by at least two native speakers. Cognitive functions assessed included psychomotor processing speed, visuomotor integration, motor speed, planning and correction, visual working memory, odour recognition and recognition of facial expressions of emotion. Computerised testing was completed on identical tablet PCs (Lenovo ThinkPad, IBM, New York) using stylus and mouse input devices; the mouse was fixed in place on a wooden platform in order to reduce participant fatigue and improve ergonomics. Since tests were administered across different countries, all instructions for French- and Dutch-

language administrations went through an extensive local translation process from the original English instructions. Brief descriptions of the tests are as follows:

i) The *Symbol Digit Modalities Test (SDMT)*[6] is a paper-and-pencil test of psychomotor speed and working memory. Participants view a 'key' at the top of the page containing symbols paired with numbers. The remainder of the page displays rows of symbols and the participant has 90 seconds to write the corresponding number that matches each symbol.

ii) Processing speed was assessed using the *Stroop Test Word Reading* condition, which is a subset of the Stroop Colour Word Interference Test.[7] Participants had 45 seconds to read aloud the words red, blue and green, printed in a series of rows.

iii) Visual attention and task switching were assessed using the *Trail Making test*,[8] which consists of 25 circles on a standard sheet of paper. For Trails A, participants were required to connect, as quickly as possible, circles containing numbers in ascending numerical order. For Trails B, participants were to connect, as quickly as possible, circles containing numbers and letters, alternating between numbers and letters in ascending order (e.g., 1, A, 2, B, 3, C, etc.).

iv) To assess visuomotor integration and motor planning, we used the *Circle-Tracing task*.[9, 10] Participants viewed a white annulus, 5 mm thick, against a grey background (i.e., a white circle with outer diameter 92.5 mm and inner diameter 87.5 mm). In each of two conditions, the participant traced the annulus by moving a stylus in a circular motion on the tablet screen. In the direct condition, the participant could directly view progress on the tablet. In the indirect condition, the participant's tracing arm and the tablet were hidden from view by a drape and progress was displayed on an ancillary screen.[10] Each condition consisted of three trials of 45-seconds each, and the order of conditions administered was counterbalanced across participants.

v) *Emotion recognition* of facial expressions of emotions was examined using computerised presentations of photographs[11] depicting six basic emotions or a neutral expression as in Johnson et al.[12] Participants were asked to indicate the emotion expressed in each photograph by selecting from the words fear, disgust, happy, sad, surprise, angry and neutral (10 stimuli per emotion).

vi) Psychomotor function was assessed in a *Paced Tapping test* using a fast and a slow condition as in Stout *et al.,*[5] and Tabrizi *et al.,*.[2] Participants tapped on left and right mouse buttons, alternating between thumbs, at 1.8 Hz (slow condition) or 3.0 Hz (fast condition). They first listened to a tone presented at the desired tapping rate, and then began tapping to the tone. After 11 taps with the tone, the repetition of the tone was discontinued, and participants attempted to continue tapping at the same rate until the end of the trial (31 taps later). Four

trials of 1.8 Hz tapping and four trials of 3 Hz tapping were completed, and whether the fast trials or the slow trials were administered first was randomised across participants.

vii) The *Serials 2's Speeded Tapping task* required participants to tap, as quickly as possible, for 10 seconds with their non-dominant index finger on a computer mouse button, whilst counting backwards aloud by 2's. The task consisted of 5 trials with each trial commencing with a different number (99, 98, 97, 96 and 95) to minimize practice effects.

viii) Visual working memory was assessed using the *Spot the Change task* that was developed by our group based on Cowan *et al*.[13] An array of coloured squares was presented for 250 ms, and, after a 1000 ms pause, a similar array was presented with one of the squares encircled. Participants were given 8000 ms to decide whether the encircled square had changed colour as compared to the same positioned square in the first array by responding 'same' (dominant thumb) or 'different' (non-dominant thumb) on the mouse buttons. A set size of 5 (i.e., number of squares presented on the screen) was used over the three data collection visits. Set size 3 was omitted after the first visit due to ceiling effects, and set size 7 was only introduced at the second visit. The design therefore provided longitudinal (i.e., 12- and 24-month) data for set size 5 only, which also presented the most symmetric distribution. The set size 5 portion of the task consisted of 32 trials.

ix) Odour recognition was assessed with a 20-item version of the *University of Pennsylvania Smell Identification Test (UPSIT; Sensonics, Inc., Haddon Heights, NJ)*[14]. Participants were required to 'scratch and sniff' 20 odour-patches and to identify the correct odour from a multiple-choice response set.

Statistical analysis

We used the standard deviation of change in the disease group (rather than pooling this with the standard deviation in controls) so that the effect sizes can be used to calculate necessary sample sizes for future clinical trials. In such sample size calculations it is commonly assumed that treatments will potentially alter mean rates of change without altering variability. Therefore, the standard deviations in HD cases are the ones of primary interest.

Considerations for the statistical model: Change in performance typically depends on the length of time between measures, and we have assumed that disease-related changes are linear over a 24 month period. Because cognitive tests are subject to improvements due to practice, and these improvements are largest from the first exposure to the second exposure, we expected practice-related change in cognitive measures to be larger over the first year than over the second year. Thus, in our statistical models, we included both time from baseline, and an indicator variable

that allowed the change between baseline and 12 months to differ from that between 12 and 24 months (over and above any effect of time interval). Because we wanted to allow the linear effect of time, differential practice effects, and average performance over all three time periods, to vary according to HD status (control, premanifest HD or early HD), we fitted GLS models with group specific intercepts, group specific slopes and group specific practice effect indicators. All analyses additionally controlled for age, sex, educational level and study site, as well as their interactions with time and practice effects. GLS models were selected because they allow for correlations in measurements from the same participant. Because we did not wish to constrain the residual variances for different visits or the covariances between pairs of visits in any way, we specified an unstructured covariance matrix, also allowing residual variances and covariances to differ according to HD group.

References

1. Warner JP, Barron LH, Brock DJH. A new polymerase chain-reaction (PCR) assay for the trinucleotide repeat that is unstable and expanded on Huntingtons-disease chromosomes. *Mol Cell Probes* 1993;**7**:235-9.
2. Tabrizi SJ, Langbehn DR, Leavitt BR, et al. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *Lancet Neurol* 2009;**8**:791-801.
3. Paulsen JS, Hayden MR, Stout JC, et al. Preparing for preventive clinical trials: The Predict-HD study. *Arch Neurol* 2006;**63**:883-90.
4. Paulsen JS, Langbehn DR, Stout JC, et al. Detection of Huntington's disease decades before diagnosis: The Predict-HD study. *J Neurol Neurosurg Psychiatry* 2008;**79**:874-80.
5. Stout JC, Paulsen J, Queller S, et al. Neurocognitive signs in prodromal Huntington disease. *Neuropsychology* 2011;**25**:1-14.
6. Smith A. Symbol Digits Modalities Test. Los Angeles: Western Psychological Services; 1991.
7. Stroop JR. Studies of interference in serial verbal reactions. *J Exp Psychol Gen* 1935;**18**:643-62.

8. Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. *Percept Mot Skills* 1958;**8**:271-6.

9. Lemay M, Fimbel E, Beuter A, Chouinard S, Richer F. Sensorimotor mapping affects movement correction defiticts in early Huntington's disease. *Exp Brain Res* 2005;**165**:454-60.

10. Say MJ, Jones R, Scahill RI, et al. Visuomotor integration deficits preced clinical onset in Huntington's disease. *Neuropsychologia* 2011;**49**:264-70.

11. Ekman P, Friesen WV. Measuring facial movement. *Environmental Psychology and Nonverbal Behavior* 1976;**1**:56-75.

12. Johnson SA, Stout JC, Solomon AC, et al. Beyond disgust: Impaired recognition of negative emotions prior to diagnosis in Huntington's disease. *Brain* 2007;**130**:1732-44.

13. Cowan N, Elliott EM, Saults JS, et al. On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cogn Psychol* 2005;**51**:42-100.

14. Doty RL, Shaman P, Kimmelman CP, Dann MS. University of Pennsylvania Smell Identification Test: A rapid quantitative olfactory function test for the clinic. *Laryngoscope* 1984;**94**:176-8.