Supplemental Data

Group Comparison of Baseline and Longitudinal Change - Single-response LMER

For each of participant and companion Symptom Checklist 90-Revised (SCL-90-R) global indices and primary symptom dimensions, three linear mixed effects regression (LMER) models were fitted separately to test for group intercept (baseline) and slope (longitudinal) differences. The three LMER models are:

Model 1 (Null model)

$$y_{ij} = (b_1 + \beta_{1i}) + (b_2 + \beta_{2i})dur_{ij} + \varepsilon_{ij}$$

Model 2 (Intercept Model)

$$y_{ij} = (b_1 + \beta_{1i}) + (b_2 + \beta_{2i})dur_{ij} + b_3 low_i + b_4 med_i + b_5 high_i + \varepsilon_{ij}$$

Model 3 (Intercept and Slope Model)

$$y_{ij} = (b_1 + \beta_{1i}) + (b_2 + \beta_{2i})dur_{ij} + b_3 low_i + b_4 med_i + b_5 high_{ii} \\ + b_6 low_i \cdot dur_{ij} + b_7 med_i \cdot dur_{ij} + b_8 high_i \cdot dur_{ij} + \varepsilon_{ij}$$

where y_{ij} is the SCL-90-R score for person i at time j; β_{1i} is the random intercept, β_{2i} is the random slope, and ε_{ij} is random error. It was assumed that two random effects follow a multivariate normal distribution with zero mean vector and unstructured covariance matrix; error terms follow a normal distribution with zero mean; and random effects and error terms are independent of each other. All models also included gender, years of education, and age at entry as covariates. The three models were evaluated using a scaling of Akaike's information criteria (AIC) corrected for small-sample bias (AICc). To rank the models, two scalings of AICc were computed: the difference in AICc (dAICc) and the AICc weight (wAICc). The dAICc was computed as the difference in AICc values between each model and the model with the lowest AICc, with smaller values indicating better fit (the best fitting model has dAICc=0 and all other models have dAICc>0). The wAICc is a probability scaling of all the AICc values (0 ≤

wAICc ≤ 1), with values closer to one indicating better fit. If Model 2 or 3 were the best fitting, then baseline and longitudinal differences between controls and each of the gene-expanded groups were reported (Table 3).

Participant and Companion Comparison - Multi-response LMER

In the multi-response LMER analysis (1), the focus is on the discrepancy between participant and companion ratings in each group over time. This discrepancy can be captured by the participant and companion slope difference. To test the slope difference between the participants and companions in each group, the participant and companion trajectories were fitted simultaneously accounting for the correlation between the participant and companion repeated measures on the same individual. The multi-response LMER was fitted stacking participant and companion data. In this stacked data, we added a new variable ‘trait’ which indicates whether the observation is from the participants or companions. Based on the trait, two indicator variables were created: (1) traitP (1 if trait=participant and 0 otherwise) and (2) traitC (1 if trait=companion and 0 otherwise). Then, these two indicator variables were used to create interactions with every independent variable in the model. In this way, stacked design matrices for the fixed and random effects can be built and we can estimate separate participant and companion fixed and random effects. The model with the group slope effect is as follows (covariates are excluded here for the simplicity but they can also be included in the same manner as groups):

For the outcome and fixed and random effects variables, participant (P) and companion (C) data were stacked as follows: $Y^{P,C} = \begin{pmatrix} Y_{ij}^{P} \\ Y_{ij}^{C} \end{pmatrix}$

$$X^{P,C} = \begin{pmatrix} cont_i^P & low_i^P & med_i^P & high_i^P & cont_i^P \cdot dur_{ij}^P & low_i^P \cdot dur_{ij}^P & med_i^P \cdot dur_{ij}^P & high_i^P \cdot dur_{ij}^P \\ cont_i^C & low_i^C & med_i^C & high_i^C & cont_i^C \cdot dur_{ij}^C & low_i^C \cdot dur_{ij}^C & med_i^C \cdot dur_{ij}^C & high_i^C \cdot dur_{ij}^C \end{pmatrix}$$

$$Z^{P,C} = \begin{pmatrix} 1^P & dur_{ij}^P \\ 1^C & dur_{ij}^C \end{pmatrix}$$

$Y^{P,C}$ is the stacked vector for the outcome variable; $X^{P,C}$ and $Z^{P,C}$ is the stacked matrix for fixed and random effect variables, respectively. Groups were coded as four dummy variables: cont, low, med, and high. Then, for each column of $X^{P,C}$ and $Z^{P,C}$, two indicator variables, traitP and traitC were multiplied separately:

$$X^{P,C}\cdot(trait=P)=\begin{pmatrix} cont_i^P & low_i^P & med_i^P & high_i^P & cont_i^P\cdot dur_{ij}^P & low_i^P\cdot dur_{ij}^P & med_i^P\cdot dur_{ij}^P & high_i^P\cdot dur_{ij}^P \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$
$$X^{P,C}\cdot(trait=C)=\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ cont_i^C & low_i^C & med_i^C & high_i^C & cont_i^C\cdot dur_{ij}^C & low_i^C\cdot dur_{ij}^C & med_i^C\cdot dur_{ij}^C & high_i^C\cdot dur_{ij}^C \end{pmatrix}$$

$$Z^{P,C}\cdot(trait=P)=\begin{pmatrix} 1^P & dur_{ij}^P \\ 0 & 0 \end{pmatrix},\ Z^{P,C}\cdot(trait=C)=\begin{pmatrix} 0 & 0 \\ 1^C & dur_{ij}^C \end{pmatrix}$$

Then,

$$Y^{P,C}=\left[X^{P,C}\cdot(trait=P),\ X^{P,C}\cdot(trait=C)\right]\begin{bmatrix} b^P \\ b^C \end{bmatrix}+\left[Z^{P,C}\cdot(trait=P),\ Z^{P,C}\cdot(trait=C)\right]\begin{bmatrix} \beta^P \\ \beta^C \end{bmatrix}+e^{P,C}$$

$$\begin{aligned} y_{ij}^{trait} &= b_1^{trait}cont_i^{trait}+b_2^{trait}low_i^{trait}+b_3^{trait}med_i^{trait}+b_4^{trait}high_i^{trait} \\ &\quad +b_5^{trait}cont_i^{trait}\cdot dur_{ij}^{trait}+b_6^{trait}low_i^{trait}\cdot dur_{ij}^{trait}+b_7^{trait}med_i^{trait}\cdot dur_{ij}^{trait}+b_8^{trait}high_i^{trait}\cdot dur_{ij}^{trait} \\ &\quad +\beta_{1i}^{trait}+\beta_{2i}^{trait}\cdot dur_{ij}^{trait}+e_{ij}^{trait} \end{aligned}$$

For any k^{th} fixed effect in the above model, we observe two separate parameters for participant (b_k^P) and companion (b_k^C).The participant and companion random effects for intercept and the duration (dur) follow multivariate normal distribution, $(\beta_{1i}^P,\beta_{2i}^P,\beta_{1i}^C,\beta_{2i}^C)^T \sim MVN(0,\Sigma)$. The error terms follow normal distribution: $e^{P,C} \sim N(0,\sigma_e^2)$.

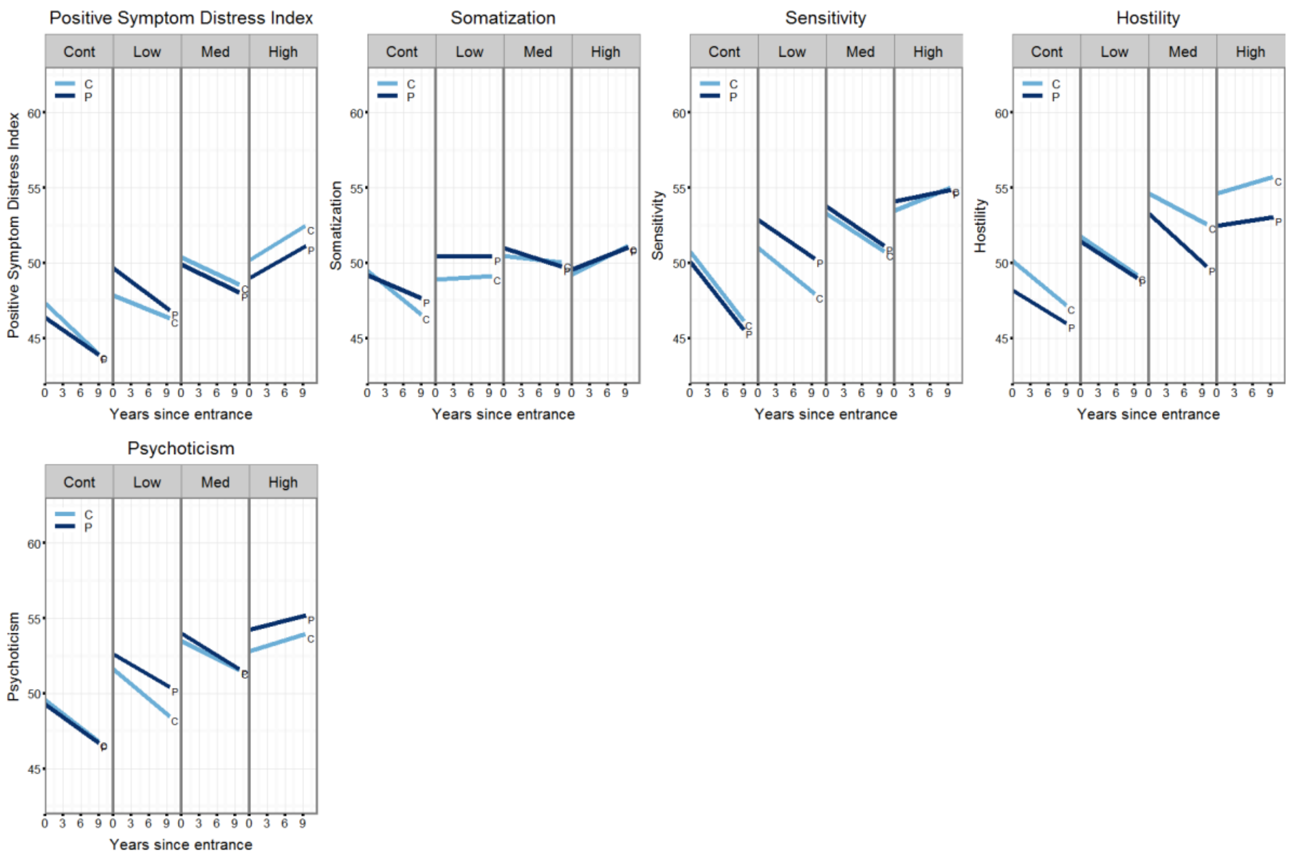
As each of the four groups (control, low, medium, and high) has either a slope discrepancy or no discrepancy between participants and companions, there were 16 possible models (Table A.1). As shown in the table A.1, Model 1 is the simplest model with no slope discrepancy in any of the groups and Model 16 is the most complex model with slope discrepancies in all groups. All models were adjusted for gender, years of education, and age at entry. These models were evaluated using AICc weights (wAICc). The relative importance of the slope discrepancy of each group was quantified by the sum of the weights over all the models with unequal group slopes (2). For example, the relative importance for the high group was assessed by summing the wAICc values over 8 models where the high group with a slope discrepancy appeared (Model 5, 8, 10, 11, 13, 14, 15, and 16). The weights close to 1 indicate higher relative importance. Model-averaged coefficients were computed by averaging model parameters over all models after multiplying the weight (wAICc) of the model and the estimated parameters for the given model (3, shows how to calculate $\tilde{\tilde{\beta}}$).

TABLE S1. 16 multi-response LMER models testing a discrepancy between the slopes of participants (P) and companions (C) ratings in each group

Model	**Group with Discrepancy**	**Group with Unequal Slopes (P slope ≠ C slope)**			
		Control	Low	Medium	High
1	None	0	0	0	0
2	Cont	1	0	0	0
3	Low	0	1	0	0
4	Med	0	0	1	0
5	High	0	0	0	1
6	Cont, Low	1	1	0	0
7	Cont, Med	1	0	1	0
8	Cont, High	1	0	0	1
9	Low, Med	0	1	1	0
10	Low, High	0	1	0	1
11	Med, High	0	0	1	1
12	Cont, Low, Med	1	1	1	0
13	Cont, Low, High	1	1	0	1
14	Cont, Med, High	1	0	1	1
15	Low, Med, High	0	1	1	1
16	Cont, Low, Med, High	1	1	1	1

Note. P = Participant; C = Companion; In column 3 to 6, 0 indicates equal P and C slopes were estimated for that group whereas 1 indicates unequal P and C slopes were estimated. All models included gender, years of education, and age at entry as covariates.

FIGURE S1. Fitted LMER curves by group for participant and companion SCL-90-R ratings. All model coefficients were estimated adjusting for gender, years of education, and age at entry. The plots show the SCL90 score as a function of duration, person (participant or companion) and group.



Supplemental Data References

1. Long, JD: Longitudinal data analysis for the behavioral sciences using R. Thousand Oaks, CA, Sage Publications, Inc., 2012, ch. 13.
2. Kenneth P, Anderson DR: Model selection and multi-model inference: a practical information-theoretic approach. New York, NY, Springer-Verlag New York LLC, 2002, ch. 4 pp 167-169.
3. Kenneth P, Anderson DR: Model selection and multi-model inference: a practical information-theoretic approach. New York, NY, Springer-Verlag New York LLC, 2002, ch. 4 pp 151-153.

PREDICT-HD Investigators, Coordinators, Motor Raters, Cognitive Raters

Isabella De Soriano, Courtney Shadrick, and Amanda Miller (University of Iowa, Iowa City, Iowa, USA); Edmond Chiu, Joy Preston, Anita Goh, Stephanie Antonopoulos, and Samantha Loi (St. Vincent's Hospital, University of Melbourne, Kew, Victoria, Australia); Phyllis Chua and Angela Komiti (University of Melbourne, Royal Melbourne Hospital, Melbourne, Victoria, Australia); Lynn Raymond, Joji Decolongon, Mannie Fan, and Allison Coleman (University of British Columbia, Vancouver); Christopher A. Ross, Mark Varvaris, Maryjane Ong, and Nadine Yoritomo (Johns Hopkins University, Baltimore); William M. Mallonee and Greg Suter (Hereditary Neurological Disease Centre, Wichita, Kan.); Ali Samii, Emily P. Freney, and Alma Macaraeg (University of Washington and VA Puget Sound Health Care System, Seattle); Randi Jones, Cathy Wood-Siverio, and Stewart A. Factor (Emory University School of Medicine, Atlanta); Roger A. Barker, Sarah Mason, and Natalie Valle Guzman (John van Geest Centre for Brain Repair, Cambridge, U.K.); Elizabeth McCusker, Jane Griffith, Clement Loy, Jillian McMillan, and David Gunn (Westmead Hospital, Sydney, Australia); Michael Orth, Sigurd Süßmuth, Katrin Barth, Sonja Trautmann, Daniela Schwenk, and Carolin Eschenbach (University of Ulm, Ulm, Germany); Kimberly Quaid, Melissa Wesson, and Joanne Wojcieszek (Indiana University School of Medicine, Indianapolis); Mark Guttman, Alanna Sheinberg, Albie Law, and Irita Karmalkar (Centre for Addiction and Mental Health, University of Toronto, Ontario); Susan Perlman and Brian Clemente (UCLA Medical Center, Los Angeles); Michael D. Geschwind, Sharon Sha, Joseph Winer, and Gabriela Satris (University of California, San Francisco); Tom Warner and Maggie Burrows (National Hospital for Neurology and Neurosurgery, London); Anne Rosser, Kathy Price, and Sarah Hunt (Cardiff University, Cardiff, U.K.); Frederick Marshall, Amy Chesire, Mary Wodarski, and Charlyne Hickey (University of

Rochester, Rochester, New York); Peter Panegyres, Joseph Lee, Maria Tedesco, and Brenton Maxwell (Neurosciences Unit, Graylands, Selby-Lemnos and Special Care Health Services, Perth, Australia); Joel Perlmutter, Stacey Barton, and Shineeka Smith (Washington University, St. Louis); Zosia Miedzybrodzka, Daniela Rae, Vivien Vaughan, and Mariella D'Alessandro (Clinical Genetics Centre, Aberdeen, U.K.); David Craufurd, Judith Bek, and Elizabeth Howard (University of Manchester, Manchester, U.K.); Pietro Mazzoni, Karen Marder, and Paula Wasserman (Columbia University Medical Center, New York); Rajeev Kumar, Diane Erickson, Christina Reeves, and Breanna Nickels (Colorado Neurological Institute, Englewood); Vicki Wheelock, Lisa Kjer, Amanda Martin, and Sarah Farias (University of California, Davis, Sacramento); Wayne Martin, Oksana Suchowersky, Pamela King, Marguerite Wieler, and Satwinder Sran (University of Alberta, Edmonton, Canada); and Anwar Ahmed, Stephen Rao, Christine Reece, Alex Bura, and Lyla Mourany (Cleveland Clinic Foundation, Cleveland).

Executive Committee: Principal Investigator Jane S. Paulsen, Jeffrey D. Long, Hans J. Johnson, Thomas Brashers-Krug, Phil Danzer, Amanda Miller, H. Jeremy Bockholt, and Kelsey Montross.

Scientific Consultants: Deborah Harrington (University of California, San Diego); Holly Westervelt (Rhode Island Hospital/Alpert Medical School of Brown University); Elizabeth Aylward (Seattle Children's Research Institute); Stephen Rao (Cleveland Clinic); David J. Moser, Janet Williams, Nancy Downing, Vincent A. Magnotta, Hans J. Johnson, Thomas Brashers-Krug, Jatin Vaidya, Daniel O'Leary, and Eun Young Kim (University of Iowa).

Core Sections

Biostatistics: Jeffrey D. Long, Ji-In Kim, Spencer Lourens (University of Iowa); Ying Zhang and Wenjing Lu (University of Indiana).

Ethics: Cheryl Erwin (Texas Tech University Health Sciences Center); Thomas Brashers-Krug, Janet Williams (University of Iowa); and Martha Nance (University of Minnesota).

Biomedical Informatics: H. Jeremy Bockholt, Jason Evans, and Roland Zschiegner (University of Iowa).