Prediction of the age of onset of Huntington Disease using a Mixed Linear Model in a Peruvian Cohort

D. Cubas-Montecino[1], M. Cornejo-Olivas [1,2], P. Mazzetti [1,3], D. Veliz-Otani[1,4,5,6]

(1)Neurogenetics Research Center, Instituto Nacional de Ciencias Neurologicas, Lima, Peru; (2) Center for Global Health, Universidad Peruana Cayetano Heredia, Lima, Peru; (3) School of Medicine, Universidad Nacional Mayor de San Marcos, Lima, Peru; (4) Fogarty Northern Pacific Global Health Fellows Program ; (5) Fogarty Interdisciplinary Cerebrovascular Diseases Training Program in South America; (6) School of Medicine, University of Maryland Baltimore. Baltimore, Maryland.

Background

- The number of CAG repeats of the expanded allele of the *HTT* gene explains ~50 to 69% of the observed phenotypic variance of the age at onset in HD.
- Heritability analyses estimate that unidentified genetic modifiers explain up to 38% of the remaining variance.
- We hypothesized that pedigree information can increase the accuracy of age of onset predictions by accounting part of unknown shared genetic modifiers.
- We aimed to predict the age of onset using mixed linear model with pedigree information.

Methods (1)

- Genotype and phenotype data from 220 affected subjects (139 probands and 81 relatives) were included in the study.
- Data was obtained from the HD registry of the Neurogenetics Research Center at the *Instituto Nacional de Ciencias Neurológicas*, Lima, Peru.
- The log(age at onset) was predicted by mixed linear models (MLM) using CAG repeat number of mutant allele, sex and matrix of kinship coefficients as covariates to model the phenotypic covariance based on additive genetic effects.
- The prediction accuracy was performed by Variance explained cross validation (VEcv) based on Leave-one-out cross-validation (LOOCV) (Table 1).

Methods (2)

Table 1. Design for Leaving-one-out cross-validation (LOOCV) of models

VEcv approach	Training dataset (n=172)		Leave-one-out in test dataset (n=1)	Model
	Related	Unrelated	Relationship to training data	
1	0	172	Unrelated	Linear regression
2	81	91 (random subset)	Unrelated	Mixed linear model (i)
3	80	92 (random subset)	Related	Mixed linear model (ii)

Results

- The mixed linear model fit to the training dataset including relatives had a greater accuracy (VEcv=63%) than the model fit with non-relatives only (VEcv=56%) (Table 2).
- The linear regression trained with unrelated subjects had a lower accuracy (VEcv=53%) than any of the mixed linear models (Table 2).

Table 2. VEcv and Pearson Coefficient of models

Model	n[a]	n[b]	r[c]	VEcv
Linear	172	139*	0.738	53%
Mixed linear (i)	172	139*	0.760	56%
Mixed linear (ii)	172	81**	0.794	63%

a: Training dataset
b: Test dataset
c: Pearson coefficient
* Unrelated subset
** Related subset

Conclusions

- Modelling the phenotypic covariance due to shared genetic effects through a matrix of kinship coefficients increases the accuracy of age of onset model prediction models.

References

1. 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(4): 398- 403.
2. 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 U S A 1997; 94(8): 3872-6.
3. 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 U S A 2004; 101(10): 3498-503
4. LI, Jin. Assessing the accuracy of predictive models for numerical data: Not r nor r2, why not? Then what?. PloS one, 2017, vol. 12, no 8, p. e0183250.

Acknowledgements & Funding

This investigation was supported by grants from FONDECYT-CONCYTEC (098-2017-FONDECYT). We are grateful to Karina Milla, Victoria Marca, and Olimpio Ortega for lab assistance. Research reported in this publication was supported by the Fogarty International Center (FIC) of the National Institutes of Health and the National Institute of Neurological Disorders and Stroke (NINDS) under grant #D43TW009345 awarded to the Northern Pacific Global Health Fellows Program and grant #D43TW009137 awarded to the Interdisciplinary Cerebrovascular Diseases Training Program in South America.

Contact

Diana Cubas-Montecino