Machine-Learning Enabled Identification of Markers of Huntington's Disease Progression

Rahul K Das[1], Jing Tu[1], Jeanne Latourelle[1], Cecilia Washburn[1], Brian Harms[1], Iya Khalil[1], John H Warner[2], Edward J Wild[3], Cristina Sampaio[2], Amrita Mohan[2]
[1]GNS Healthcare, Cambridge, MA 02139; [2]CHDI Foundation, Princeton, NJ 08540; [3]UCL Huntington's Disease Centre, Queen Square, London WC1N 3BG

Objectives

To date, few prognostic biomarkers for Huntington's Disease (HD) progression have been validated, inhibiting the development of novel therapeutics.

Specific objective outcome measures have been proposed (Tabrizi et al. 2012) to be sensitive to subtle changes occurring at premanifest stage.

Identification of specific prognostic markers these outcomes may allow:

- Early identification of fast-progressing patients
- Identification of subpopulations with differing progression profiles
- Mechanistic insights underlying progression differences

Bayesian machine learning models of early-HD progression were created using clinical, imaging, genetic, and biomarker data from the TRACK-HD and TRACK-ON longitudinal studies. These models were used to identify patient subpopulations with differing progression rates and mechanistic insights underlying these progression differences.

Study Population and Data

DISCOVERY SET: TRACK-HD/TRACK-ON (Tabrizi et al. 2013)

- 279 Early HD (EHD), Premanifest (PM) and healthy control (HC) patients with longitudinal clinical progression outcomes and genetic data

VALIDATION SET: PREDICT-HD (Zhang et al. 2011)

- 892 Premanifest (PM) and healthy control (HC) patients with longitudinal clinical progression outcomes and genetic data

Progression was evaluated for three different time periods resulting in different study populations

	TRACK-HD/TRACK-ON Training Set		PREDICT-HD Validation Set	
Time Frame	Visits	Patients (% HC, % PM)	Visits	Patients (% HC, % PM)
12-Month Progression	1284	279 (32%,37%)	2706	892 (21%,79%)
36-Month Progression	734	276 (32%,37%)	1110	540 (14%,86%)
60-Month Progression	301	154 (46%,54%)	234	213 (14%,86%)

STUDY OUTCOME MEASURES:

Rate of clinical progression of three clinical metrics

- UHDRS Total Motor Score (TMS)
- Symbol Digit Modalities Test (SDMT): Identified as sensitive markers of disease progression in Early-HD patients (Tabrizi et al. 2012)
- Composite Unified Huntington Disease Rating Scale (cUHDRS): A composite measure of motor, cognitive, and global functional decline that provides an improved measure of clinical progression in HD (Schobel et al. 2017)

Compared effectiveness of predictors over different time courses

- Short-Term: Predicting change in outcome in 12 months
- Longer-Term: Predicting change in outcome in 36 and 60 months

Potential predictors included from demographic (N=8), clinical (N=48 - 93), neurological imaging (N=3), known snps (N=77), summarized genotype scores (N = 843), and plasma biomarkers (N=2).

Methods

GNS Healthcare's proprietary machine learning platform, Reverse Engineering and Forward Simulation (REFS™) was used to build predictive models. Selection of a single model underestimates prediction error, thus REFS learns an ensemble of the most probable models (N=128) given the data.
Ensemble constructed via Monte Carlo sampling of the posterior model landscape.

- Model additions/ subtractions scored based on a maximum entropy structural prior with complexity also penalized by the Bayesian Information Criterion.
- Linear, additive, quadratic, and cubic terms allowed in order to accommodate non-linear effects and sub-populations.
- Confidence of a given relationship X->Y determined by frequency among ensemble.

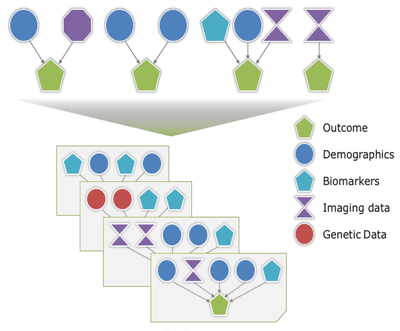


Figure 1. Visualization of REFS™ enumeration of model fragments and reverse-engineering of prediction model ensemble.

- Linear mixed effect model was used to account for longitudinal data
- Model performance was evaluated using marginal R^2: the proportion of variance explained by the fixed factor(s) alone
- Gene scores were calculated using Rasch modeling (Wang et al., 2015). These scores estimated the latent trait of a defined gene deduced by its SNP profile.

Key clinical and genetic predictors

Motor Score
- Short and long-term motor progression
- Longer term cognitive progression

CAP score
- Short and long-term motor progression
- 12- and 36-month change in cUHDRS

Plasma NfL concentration
- Long-term motor progression
- Short and long-term cognitive progression

Caudate and ventricular volumes
Long-term motor progression & change in cUHDRS

Performance in Spot-the-change test
Short-term motor & cognitive progression

Performance in Stroop word reading test
- Short-term motor progression
- Longer-term change in cUHDRS

Genetics (Rasch Scores)
- **FRA10AC1 & PIK3C2A:** long term change in cUHDRS
- **CHAC2:** 36-month change in cUHDRS

Figure 2A: Caudate volume was predictor of TMS change in EHD (TRACK)

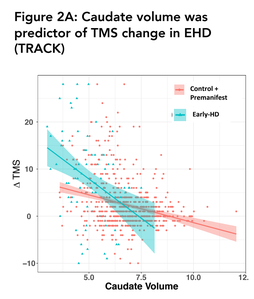


Figure 2B: Baseline NfL concentration was predictor of 36-mo. SDMT change in PM and EHD (TRACK)

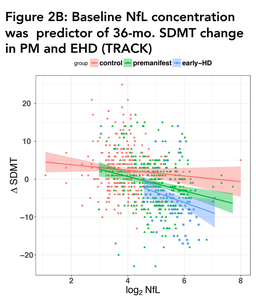
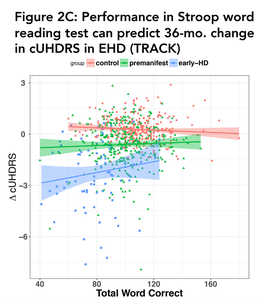


Figure 2C: Performance in Stroop word reading test can predict 36-mo. change in cUHDRS in EHD (TRACK)



Validation of novel genetic predictors

Figure 3: Distributions of 60-month change of cUHDRS for the different genotypes of SNP Chr11: 17191019A>G (rs214936) in PIK3C2A. (Left) TRACK-HD, (Right) PREDICT-HD

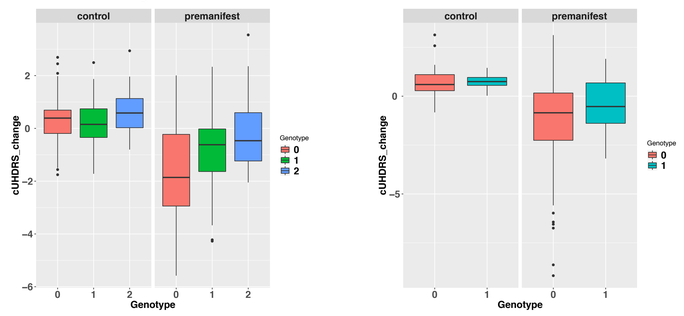
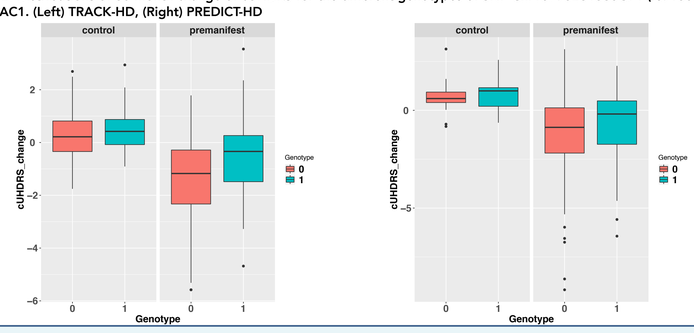


Figure 4: Distributions of 60-month change of cUHDRS for the different genotypes of SNP Chr10: 95431535C>T (rs913045) in FRA10AC1. (Left) TRACK-HD, (Right) PREDICT-HD



- Two novel loci PIK3C2A (rs214936) and FRA10AC1 (rs913045) were identified as predictive of cUHDRS progression in the TRACK-HD/ON cohorts.
- Validation of these SNPs was conducted in the PREDICT-HD cohort and statistically significant (PIK3C2A p=0.049; FRA10AC1 p= 0.04) replication was observed in both loci.

Predictive Performance

Outcome	Objective	In-sample Marginal R^2	k-fold* Cross-validation Marginal R^2
ΔTMS	12-Month	0.18	0.13
	36-Month	0.31	0.14
	60-Month	0.50	0.28
ΔSDMT	12-Month	0.12	0.09
	36-Month	0.19	0.11
	60-Month	0.26	0.13
ΔcUHDRS	12-Month	0.12	0.11
	36-Month	0.32	0.19
	60-Month	0.45	0.26

Model performance could not be properly evaluated in the PREDICT cohort due to missingness of several key predictors and differences in populations

Differentiating fast vs slow progressors

Questions:
- How well can models differentiate between "fast and "slow" progressing patients?
- How does this compare to known predictors?

Approach:
- Identify patients (gene+ only) as "fast" and "slow" progressors by dichotomizing on median
- Use model based predictions to calculate AUC for correctly identifying patients into each group
- Compare to AUC using CAP score to predict group

Figure 4: Differentiating fast vs slow progressing patients: (Left) 12-month change in TMS (TRACK-HD), (Right) 12-month change in TMS (PREDICT-HD)

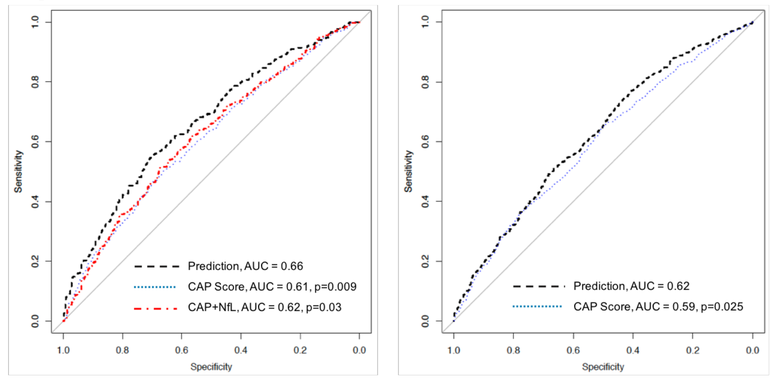
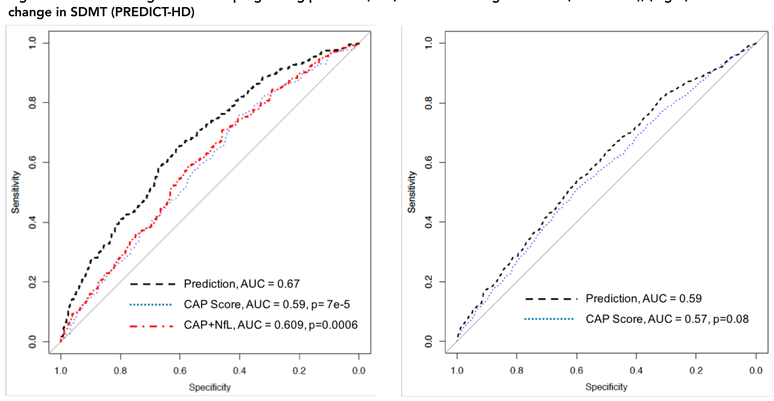


Figure 5: Differentiating fast vs slow progressing patients: (Left) 12-month change in SDMT (TRACK-HD), (Right) 12-month change in SDMT (PREDICT-HD)



- For 12-month change in TMS, model based predictions outperformed CAP score alone and CAP score + NfL in both TRACK and PREDICT.
- A similar trend was observed for 12-month change in SDMT in TRACK but not in PREDICT due to missingness of important key predictors such as, imaging markers.

Conclusions

- Novel loci in FRA10AC1 and PIK32CA genes were significantly associated with the 60-month change in cUHDRS in two independent HD cohorts, thus providing mechanistic insights underlying progression differences.
  - FRA10AC1 contains a fragile site, many of which have been previously implicated in neurodevelopmental disease (Metsu et al., 2014).
  - variants in FRA10AC1 are associated with CSF Aβ1–42 level passing the genome-wide significance threshold (Li et al., 2015).
  - shRNA knock-down of PIK3C2A suppresses mHTT aggregation (Yamanaka et al., 2014).
- Model predictions differentiated slower and faster progressing patients more effectively than CAP score alone for TMS and SDMT in both short and long term allowing early identification of fast-progressing patients, which is crucial for laying out a well-defined roadmap for treatments targeted at stages before symptom onset.
- Several key clinical predictors were identified as potential markers of subpopulations with differing progression profiles.
  - Spot-the-change test, NfL, Imaging markers (caudate and ventricular volumes), Performance in Stroop word reading test
- These findings suggest a toolkit to predict HD progression that may have utility for clinical trial programs

Acknowledgements

This work was generously supported by the CHDI Foundation.

References

Tabrizi et al. (2012) Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: analysis of 24 month observational data. Lancet Neurol. 11(1):42-53.
Tabrizi et al. (2013) Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. Lancet Neurol.12(7):637-649.
Zhang et al. (2011) Indexing Disease Progression at Study Entry with Individuals At-Risk for Huntington Disease. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet. 156, 751–763.
Lee J-M et al. (2015) Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. Cell 162:516–526.
Li QS, Parrado AR, Samtani MN, Narayan VA, Alzheimer's Disease Neuroimaging Initiative (2015) Variations in the FRA10AC1 Fragile Site and 15q21 Are Associated with Cerebrospinal Fluid Aβ1-42 Level. PLOS ONE 10:e0134000.
Metsu S, Rooms L, Rainger J, Taylor MS, Bengani H, Wilson DI, Chilamakuri CSR, Morrison H, Vandeweyer G, Reyniers E, Douglas E, Thompson G, Haan E, Gecz J, FitzPatrick DR, Kooy RF (2014) FRA2A Is a CGG Repeat Expansion Associated with Silencing of AFF3. PLOS Genet 10:e1004242.
Schobel SA, Palermo G, Auinger P, Long JD, Ma S, Khwaja OS, Trundell D, Cudkowicz M, Hersch S, Sampaio C, Dorsey ER, Leavitt BR, Kieburtz KD, Sevigny JJ, Langbehn DR, Tabrizi SJ (2017) Motor, cognitive, and functional declines contribute to a single progressive factor in early HD. Neurology 89:2495.
Wang W, Mandel J, Bouaziz J, Commenges D, Nabirotchkine S, Chumakov I, Cohen D, Guedj M, the Alzheimer's Disease Neuroimaging Initiative (2015) A Multi-Marker Genetic Association Test Based on the Rasch Model Applied to Alzheimer's Disease. PLOS ONE 10:e0138223.
Yamanaka T, Wong HK, Tosaki A, Bauer PO, Wada K, Kurosawa M, Shimogori T, Hattori N, Nukina N (2014) Large-scale RNA interference screening in mammalian cells identifies novel regulators of mutant huntingtin aggregation. PloS One 9:e93891–e93891.