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.

<table>
<thead>
<tr>
<th>Time Frame</th>
<th>TRACK-HD/TRACK-ON Training Set</th>
<th>PREDICT-HD Validation Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-Month Progression</td>
<td>Visits (%) Patients</td>
<td>Visits (%) Patients</td>
</tr>
<tr>
<td>1284</td>
<td>(32%,37%)</td>
<td>279</td>
</tr>
<tr>
<td>734</td>
<td>(32%,37%)</td>
<td>110</td>
</tr>
<tr>
<td>301</td>
<td>(15%,45%)</td>
<td>234</td>
</tr>
</tbody>
</table>

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 (UCHDRS): A composite measure of motor, cognitive, and global functional decline that provides an improved measure of clinical progression in HD (Schuff et al. 2017)

Compared effectiveness of predictors over different time courses
• Short-Term: Predicting change in outcome in 12 months
• Long-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 genotypes 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 underspecified 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® estimation of model fragments and reverse-engineering of prediction model ensemble.

PREDICTIVE PERFORMANCE
Outcome | Objective | In-sample | k-fold | Cross-validation | Marginal R² | Marginal R² |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TMS</td>
<td>12-Month</td>
<td>0.18</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36-Month</td>
<td>0.31</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60-Month</td>
<td>0.50</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASDMT</td>
<td>12-Month</td>
<td>0.12</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36-Month</td>
<td>0.19</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60-Month</td>
<td>0.26</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACHDRT</td>
<td>12-Month</td>
<td>0.12</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36-Month</td>
<td>0.19</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60-Month</td>
<td>0.26</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nGHDRT</td>
<td>12-Month</td>
<td>0.45</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

REFERENCES

CONCLUSIONS
• Novel loss in FRA10AC1 and PFK2CA genes were significantly associated with the 60-month change in CHDRT 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 (Mehta et al. 2014).
  - Variants in FRA10AC1 are associated with CSP A1 gene level passing the genome-wide significance threshold (Li et al., 2015).
  - shRNA knock-down of PFK2CA suppresses mHTT aggregation (Yamanaka et al., 2014).
• Model predictions differentiated slower and faster progressing patients more effectively than model change 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, NLS, Imaging markers (cavate 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.