Analysis and classification of transcriptome dysregulation signals over the course of Huntington's disease progression in mouse model systems.

Jonathan R. Greene[1], Marissa B. Hirst[1], John C. Obenauer[1], Paul Donovan[1], Jeff Aaronson[2], Jian Chen[2], Jim Rosinski[2]

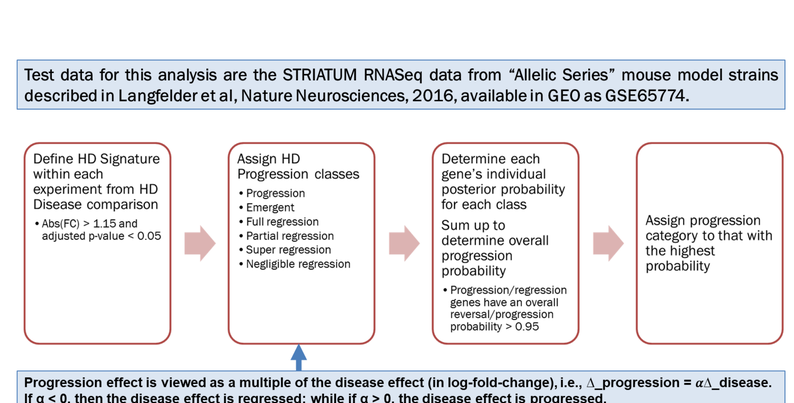
[1]Rancho BioSciences, San Diego, USA; [2]CHDI Management Inc. (the company that manages the scientific activities of CHDI Foundation, Inc.) NY, USA

Introduction

Numerous studies in animal model systems for Huntington's disease have described large gene expression effects in various tissues, especially in the striatum region of the brain. In studies involving samples across either different age groups or other disease related factors (such as Q-length) it can be of interest to characterize how these factors affect dysregulated genes or proteins over the course of disease progression. We describe here a Bayesian approach that utilizes differential gene expression results from disease and progression related contrasts to classify each gene's trajectory through the course of the disease model. For example, the method can describe genes in terms of when their dysregulation appears and the extent to which their dysregulation continues to advance or is sustained. The method is based on a previously published method to measure reversal or rescue of gene dysregulation (Marchionini, et al, JCI Insight, 2022). Examples will focus on well characterized published data sets. We will explore the implications for marrying these results to transcriptome-wide effects from disease model intervention studies.

Description of the Method

- Studies amenable to this analysis involve HD model systems that span multiple ages, Q-lengths or other disease parameters.
- The analysis paradigm is based on three differential expression contrasts:
  - **HD Signature contrast: HD Model vs Wild Type at a starting parameter value**
    - Eg: Q111 at 6M vs WT at 6M
  - **HD Progression contrast: HD Model at the next parameter value vs HD Model at the starting parameter value**
    - Eg: Q111 at 10M vs Q111 at 6M
  - **Next HD Signature contrast: HD Model at the next progression parameter value vs Wild Type at the next progression parameter value**
    - Eg: Q111 at 10M vs WT at 10M)
    - This contrast captures genes that "Emerge" as HD signature genes between 6 months and 10 months
- Disease progression probabilities and classifications are determined using a Bayesian approach originally designed for gene-by-gene "reversal/rescue/prevention" as described in Marchionini, et al, *JCI Insight*, 2022.



Difference in Differences plots such as this show how individual genes respond in
- HD Signature in Striatum (x-axis):
  - Q111 vs WT at 6M
- HD Progression in Striatum (y-axis):
  - Q111 at 10M vs Q111 at 6M

Data for this example use striatal data from the "Allelic Series" data from Langfelder, et al (2016), GSE65774

Genes in quadrants A and B represent genes whose expression effect in the HD signature PROGRESSES (effects are in the SAME direction)

Genes in quadrants C and D represent genes whose expression effect in the HD Signature REGRESSES (effects are in the OPPOSITE direction)

Genes that are colored BLUE and are close to the X-axis represent genes that NEITHER PROGRESS NOR REGRESS

Genes that are colored GREEN and are close to the Y-axis represent genes that were not part of the HD signature at 6 months but EMERGE as HD signature genes at 10 months

REGRESSION quadrants further subdivided
- GREEN shaded area – Full Regression
- ORANGE shaded area – Partial Regression
- PURPLE shaded area – Super Reggression

Individual gene Log2FCs (point) and Standard Error (error bars and ellipse) establish probability that a given gene (point) falls within a given regression category (area of the plot).

Test data for this analysis are the STRIATUM RNASeq data from "Allelic Series" mouse model strains described in Langfelder et al, Nature Neurosciences, 2016, available in GEO as GSE65774.

Progression effect is viewed as a multiple of the disease effect (in log-fold-change), i.e., Δ_progression = $\alpha\Delta$_disease.
If α < 0, then the disease effect is regressed; while if α > 0, the disease effect is progressed.
The PP method examines 5 possible cases for α, for example:
- **Progression: α > 0.2**
- **Super-regression: α < -1.3**
- **Full regression: -1.3 < α < -0.7**
- **Partial regression: -0.7 < α < -0.2**
- **Negligible regression: -0.2 < α < 0.2**

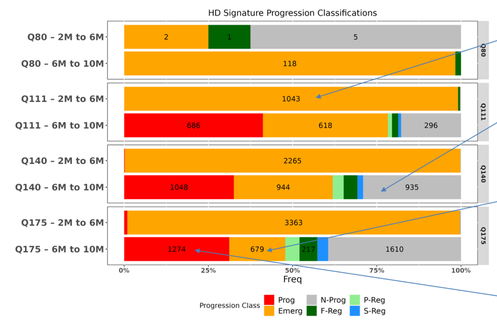
Emergent genes are those that are not part of the starting HD Signature but then appear in the HD Signature at the next disease progression parameter.

Results

Test data for this analysis are the STRIATUM RNASeq data from "Allelic Series" mouse model strains described in Langfelder et al, Nature Neurosciences, 2016, available in GEO as GSE65774.

Progression in the Allelic Series by Age across Q-lengths

Disease Progression classification gene counts for Allelic Series progression by Age at Q80, Q92, Q111, Q140, and Q175 from 2M to 6M and 6M to 10M represented as stacked percentage bar plots. X-axis indicates cumulative percentage of HD signature genes, numbers of genes in each progression class are indicated, progression classes are colored as indicated in legend.



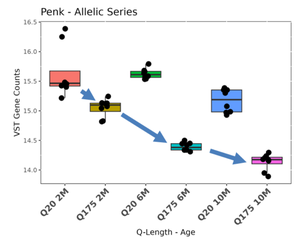
For all Q-lengths, majority of genes are Emergent ("Emerg", orange) in 2M to 6M progression.

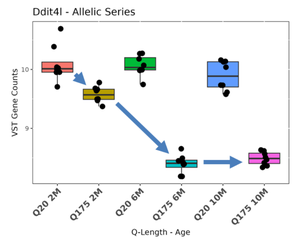
For 6M to 10M progressions, fraction of Negligible progressions ("N-Prog", grey) increases with increasing Q-length from Q111 to Q175.

For 6M to 10M progressions, fraction of Emergent genes ("Emerg", orange) decreases with increasing Q-length from Q111 to Q175.

Fraction of Progressing genes ("Prog", red) decreases slightly with increasing Q-length from Q111 – Q175.

Example single gene plots from Allelic Series progression by age analysis showing Progression and Negligible Progression classes.
Blue arrows indicate trajectory of progressive dysregulation.



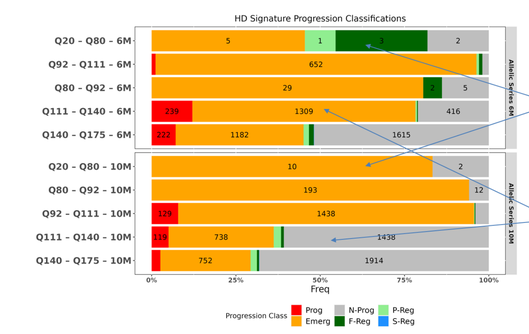


Scenario	Class	Disease LFC	Disease FDR	Aging LFC	Disease Stat	Aging Stat	Emerging Stat
Q175 – 2M to 6M	Progression	-0.64	4.11e-02	-0.66	-3.99	-11.71	-32.4
Q175 – 6M to 10M	Negligible Progression	-1.24	6.24e-226	-0.25	-32.40	-4.72	-13.2

Scenario	Class	Disease LFC	Disease FDR	Aging LFC	Disease Stat	Aging Stat	Emerging Stat
Q175 – 2M to 6M	Progression	-0.58	5.44e-03	-1.35	-4.56	-16.66	-21.93
Q175 – 6M to 10M	Negligible Progression	-1.88	3.54e-103	0.11	-21.93	1.36	-15.64

Progression in the Allelic Series, Q-length as a surrogate for Age

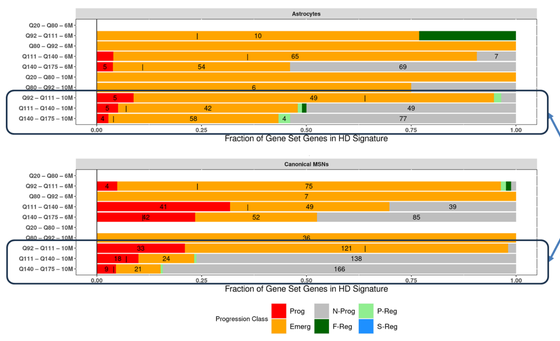
Disease Progression classification gene counts for Allelic Series progression using Q-length as a surrogate for age at 6M and 10M represented as stacked percentage bar plots. X-axis indicates cumulative percentage of HD signature genes, numbers of genes in each progression class are indicated, progression classes are colored as indicated in legend.



HD Signatures for Q20 & Q80 contain relatively small numbers of genes.

For Q92 – Q175 progression, fraction of Negligible progression genes ("N-Prog", grey) increases while fraction of Emergent genes ("Emerg", orange) decreases.

Astrocyte and Canonical MSN cell type marker genes from DropViz, disease progression classification gene counts for Allelic Series progression using Q-length as a surrogate for age at 6M and 10M represented as stacked percentage bar plots. X-axis indicates cumulative percentage of HD signature genes, numbers of genes in each progression class are indicated, progression classes are colored as indicated in legend.



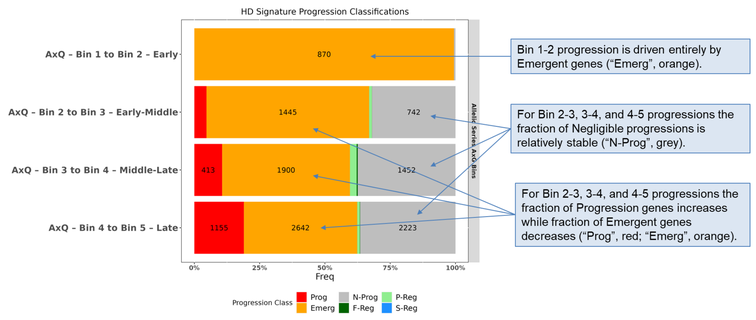
Difference observed at 10M in relative proportions of Progressions, Emergent genes, and Negligible Progressions between Astrocyte and Canonical MSN cell type markers from Q92 to Q175.

Progression in the Allelic Series using the product of Age and Q-length

- **In order to define a framework to which results from various HD intervention studies can be compared, we multiply Age * Q-Length to define an "AxQ" value**
  - **AxQ is discrete owing to fixed Age and Q-Length values in the Allelic Series**
  - **AxQ scores were binned into FIVE groups as indicated in the table below by intervals of 300**
- **HD Signature contrasts are defined as AxQ Bin X vs WT 2M**
- **Progression contrasts are defined as follows:**
  - **Bin 2 vs Bin 1 – "Early"**
  - **Bin 3 vs Bin 2 – "Early-Middle"**
  - **Bin 4 vs Bin 3 – "Middle-Late"**
  - **Bin 5 vs Bin 4 – "Late"**

	Q20	Q80	Q92	Q111	Q140	Q175
2M	AxQ 0-300, Bin 1	AxQ 0-300, Bin 1	AxQ 0-300, Bin 1	AxQ 0-300, Bin 1	AxQ 0-300, Bin 1	AxQ 300-600, Bin 2
6M	AxQ 0-300, Bin 1	AxQ 300-600, Bin 2	AxQ 300-600, Bin 2	AxQ 600-900, Bin 3	AxQ 600-900, Bin 3	AxQ 900-1200, Bin 4
10M	AxQ 0-300, Bin 1	AxQ 600-900, Bin 3	AxQ 900-1200, Bin 4	AxQ 900-1200, Bin 4	AxQ > 1200, Bin 5	AxQ > 1200, Bin 5

Disease Progression classification gene counts for Allelic Series progression using AxQ bins, represented as stacked percentage bar plots. X-axis indicates cumulative percentage of HD signature genes, numbers of genes in each progression class are indicated, progression classes are colored as indicated in legend.



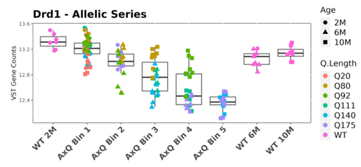
Bin 1-2 progression is driven entirely by Emergent genes ("Emerg", orange).

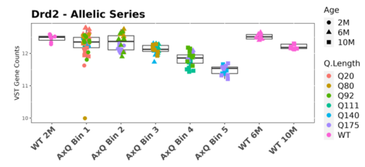
For Bin 2-3, 3-4, and 4-5 progressions the fraction of Negligible progressions is relatively stable ("N-Prog", grey).

For Bin 2-3, 3-4, and 4-5 progressions the fraction of Progression genes increases while fraction of Emergent genes decreases ("Prog", red; "Emerg", orange).

Results

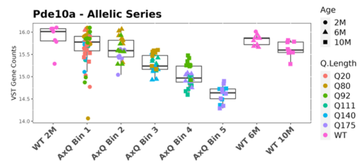
Progressive disease associated dysregulation for selected genes is apparent in declining expression from WT 2M across increasing AxQ Bins
(Key differential expression parameters are presented in embedded tables)

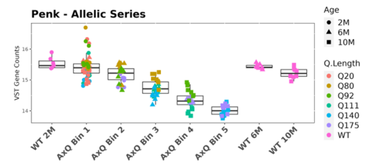




Scenario	Class	Disease LFC	Disease FDR	Aging LFC	Disease Stat	Aging Stat	Emerging Stat
AxQ – Bin 1 to Bin 2 – Early	Emergent	-0.14	7.06e-01	-0.32	-2.21	-4.06	4.07
AxQ – Bin 2 to Bin 3 – Early-Middle	Progression	-0.32	7.19e-03	-0.21	-4.07	-3.63	-5.09
AxQ – Bin 3 to Bin 4 – Middle-Late	Progression	-0.53	2.94e-05	-0.21	-5.09	-2.57	-4.63
AxQ – Bin 4 to Bin 5 – Late	Negligible Progression	-0.74	2.73e-09	-0.23	-6.63	-2.96	-17.06

Scenario	Class	Disease LFC	Disease FDR	Aging LFC	Disease Stat	Aging Stat	Emerging Stat
AxQ – Bin 1 to Bin 2 – Early	Emergent	-0.11	6.43e-01	-0.34	-1.13	-3.95	-5.81
AxQ – Bin 3 to Bin 4 – Middle-Late	Progression	-0.34	1.13e-06	-0.29	-5.81	-5.53	-7.61
AxQ – Bin 4 to Bin 5 – Late	Progression	-0.64	4.87e-12	-0.35	-7.61	-5.79	-14.78





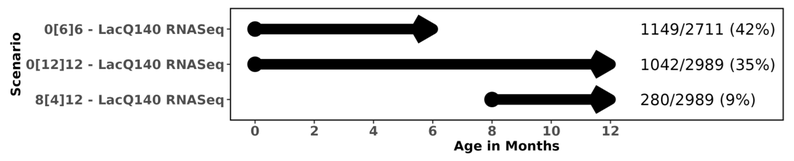
Scenario	Class	Disease LFC	Disease FDR	Aging LFC	Disease Stat	Aging Stat	Emerging Stat
AxQ – Bin 1 to Bin 2 – Early	Emergent	-0.27	1.10e-01	-0.60	-2.52	-5.09	4.61
AxQ – Bin 3 to Bin 4 – Middle-Late	Progression	-0.60	1.68e-08	-0.27	-6.61	-4.37	-8.53
AxQ – Bin 4 to Bin 5 – Late	Progression	-0.87	5.28e-15	-0.41	-8.53	-6.02	-14.33

Scenario	Class	Disease LFC	Disease FDR	Aging LFC	Disease Stat	Aging Stat	Emerging Stat
AxQ – Bin 1 to Bin 2 – Early	Emergent	-0.20	6.90e-01	-0.36	-2.26	-2.58	-3.43
AxQ – Bin 2 to Bin 3 – Early-Middle	Progression	-0.36	2.51e-02	-0.43	-3.43	-5.61	-7.20
AxQ – Bin 3 to Bin 4 – Middle-Late	Progression	-0.78	3.99e-10	-0.40	-7.20	-4.90	-10.18
AxQ – Bin 4 to Bin 5 – Late	Progression	-1.19	4.98e-21	-0.34	-10.18	-4.31	-20.44

Overlay Progression classes with Reversal/Rescue results

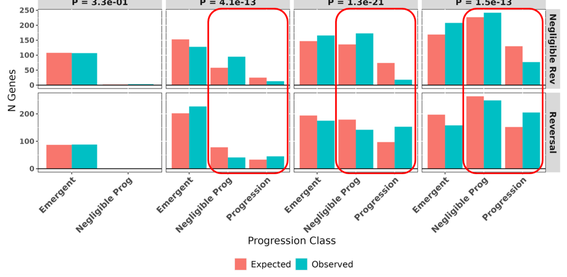
Phenotype REVERSAL/PREVENTION/RESCUE results from the LacQ140 Huntingtin lowering model described in Marchionini, et al, JCI Insight, 2022. Percentage of reversed genes is presented along with numbers for HD Signature genes and genes that are overall reversed. Start and end points represent administration of perturbation and age of harvest. Y-axis labels represent shorthand notation for the three Huntingtin lowering scenarios.

Are different Allelic Series AxQ Disease Progression categories more amenable to reversal/rescue/prevention than others?



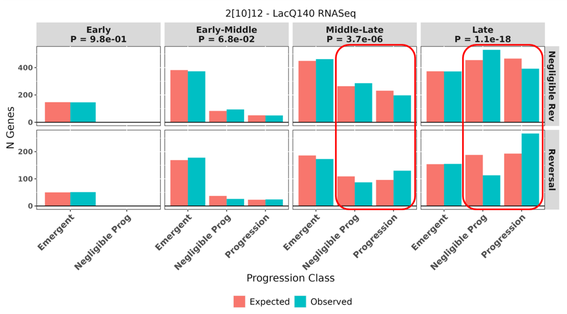
Are different Allelic Series AxQ Disease Progression categories more amenable to reversal/rescue/prevention than others in LacQ140 Huntingin lowering model?

Overlay Allelic Series AxQ progression classifications with LacQ140 reversal/rescue/prevention classifications and test using Chi-square (P-values indicated in plot headers)



For this early intervention for Huntington lowering:

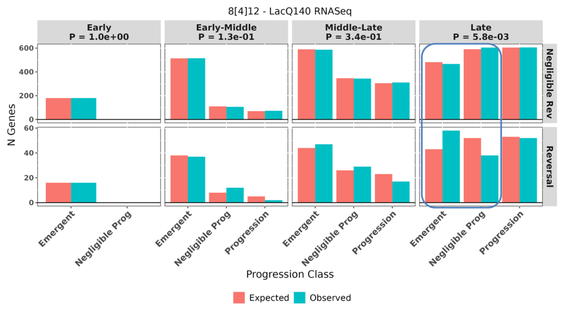
Early-Middle, Middle-Late and Late "Progression" genes are MORE likely to be reversed, Negligible Progression genes are LESS likely to be reversed. (red boxes)



For this early intervention for Huntington lowering:

Middle-Late and Late "Progression" genes are MORE likely to be reversed, Negligible Progression genes are LESS likely to be reversed. (red boxes)

Note that in the effect is less pronounced in this 12M cohort than in the 6M cohort shown above.



For this late intervention for Huntington lowering:

Overall reversal is very low and there is a preference for reversal of Emergent genes at the expense of Negligible Progression genes (blue box)

Summary and Conclusions

- Transcriptome progression signals can be described through adaptation of the posterior probability method used for reversal/rescue
- Applied across various progression parameters
  - Age
  - Q-length as a surrogated for Age
  - Age * Q-length (AxQ)
- Results can be overlayed with Reversal/Rescue calculations
  - Reversal is possible for genes in all progression classes
  - For early intervention: trend towards more effective reversal/rescue/prevention of Middle-Late and Late Progressors as compared to established HD genes (Negligible Progressions) and Early and Early-Middle Progressors (this result has been observed in other studies not shown here).
  - For late intervention: trend towards more effective reversal/rescue/prevention of Late Emergent genes over other classes