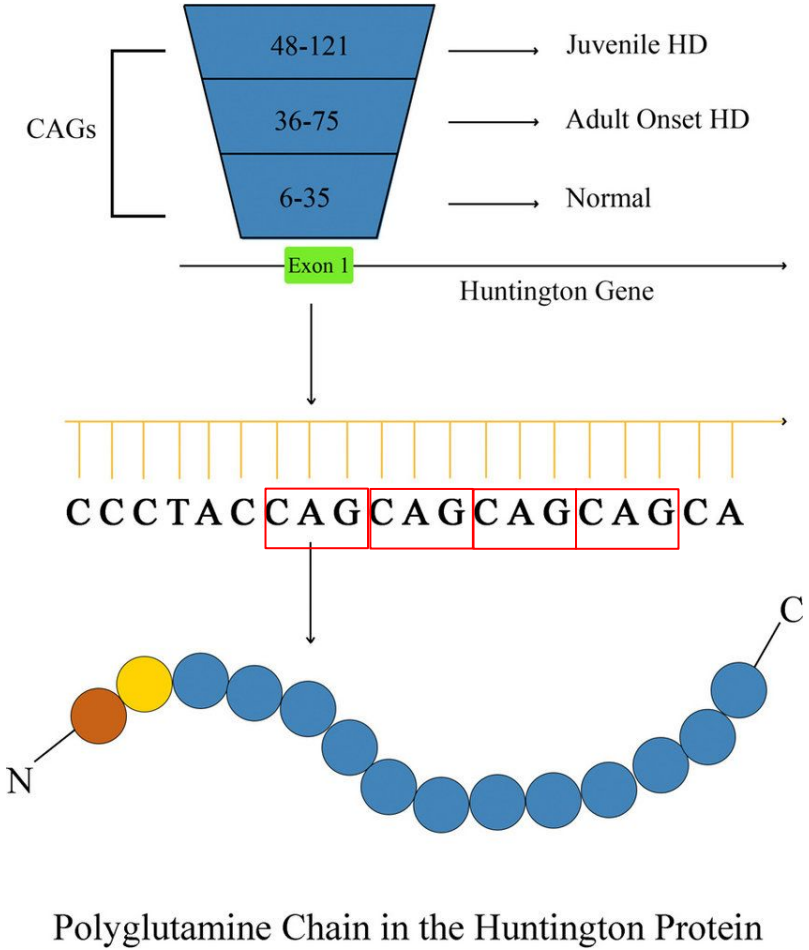
Updated Models for Somatic CAG Repeat Expansion in Huntington's Disease

by Steven Tan, mentored by Bob Handsaker and Seva Kashin

Project Conducted in the Steve McCarroll’s Lab, Harvard Medical School

Huntington’s Disease

- Huntington’s Disease (HD) is a fatal, inherited neurodegenerative disease.
- Caused by inheriting high number of CAG repeats (36+) in the Huntingtin (*HTT*) gene.
- Patients with the disease exhibit somatic expansion, where the number of repeats expands individually in each brain cell.
- High CAG lengths can be toxic and lead to neuronal cell death.



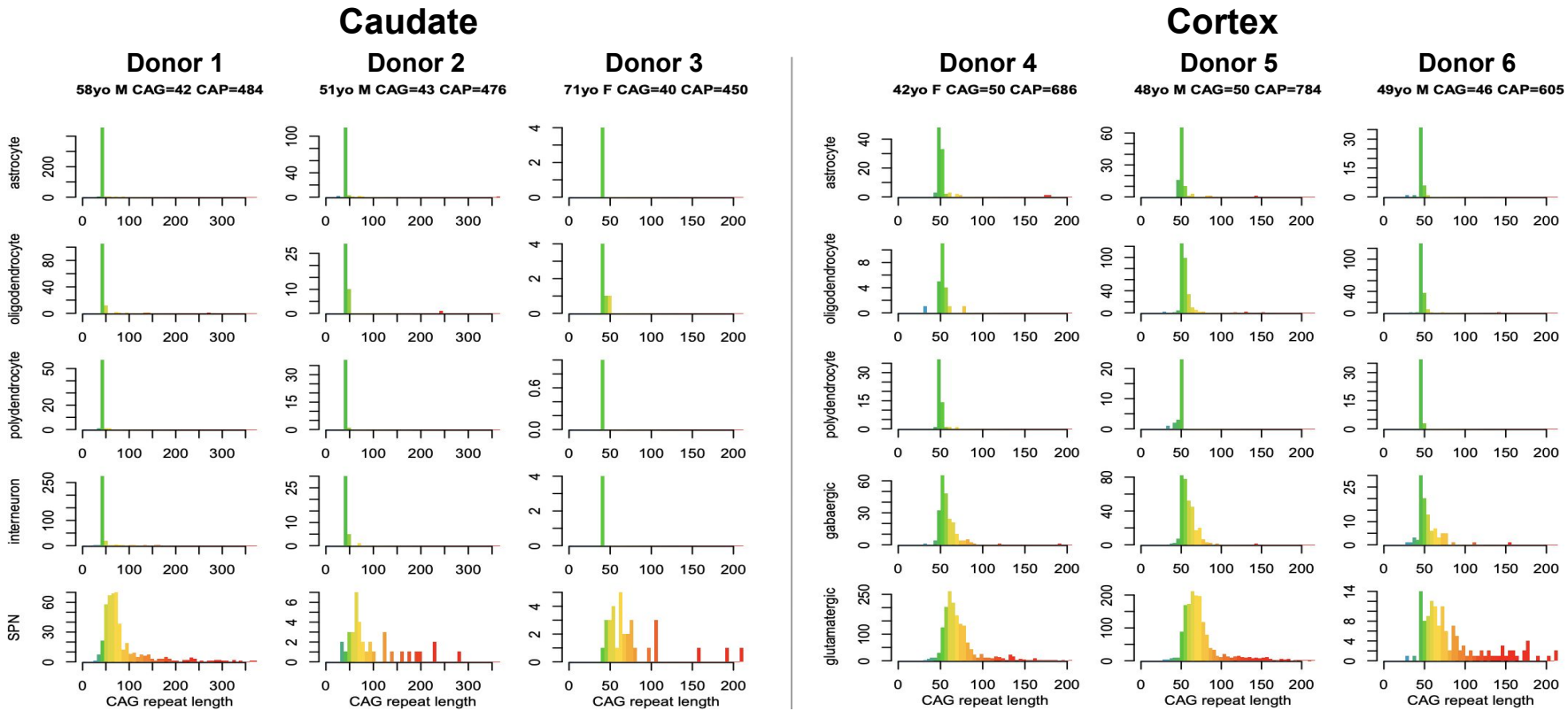
Polyglutamine Chain in the Huntington Protein

Figure from: https://pubmed.ncbi.nlm.nih.gov/32811395/

Research Goals

- Somatic expansion is increasingly thought to be the driver of disease progression.
- Our goal is to create a statistical model that simulates somatic CAG repeat expansion.
- Such a model could help us better understand mechanisms underlying somatic expansion.
    - Important for developing therapeutics treating HD and for designing clinical trials.

Data Collected from Post-Mortem Brain Samples of HD Patients



Collected by the McCarroll Lab in collaboration with the McLean Brain Bank

Modeling Somatic Expansion

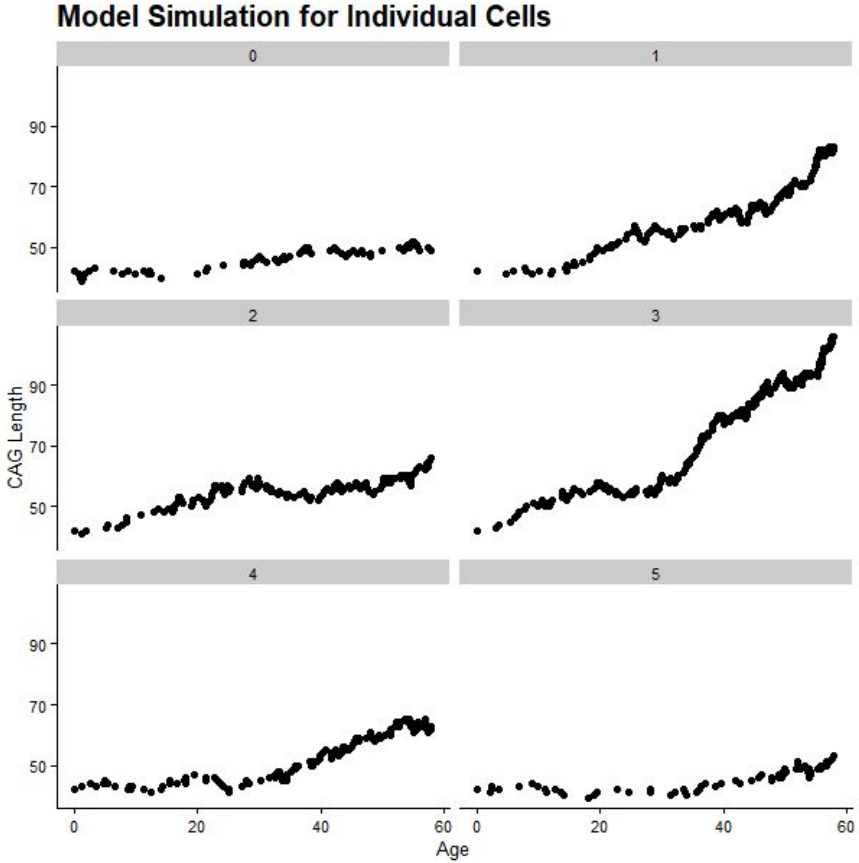
The model was inspired by the work of Dr. John Warner and previous models applied for other repeat expansion diseases [1].

The generative model works as follows:

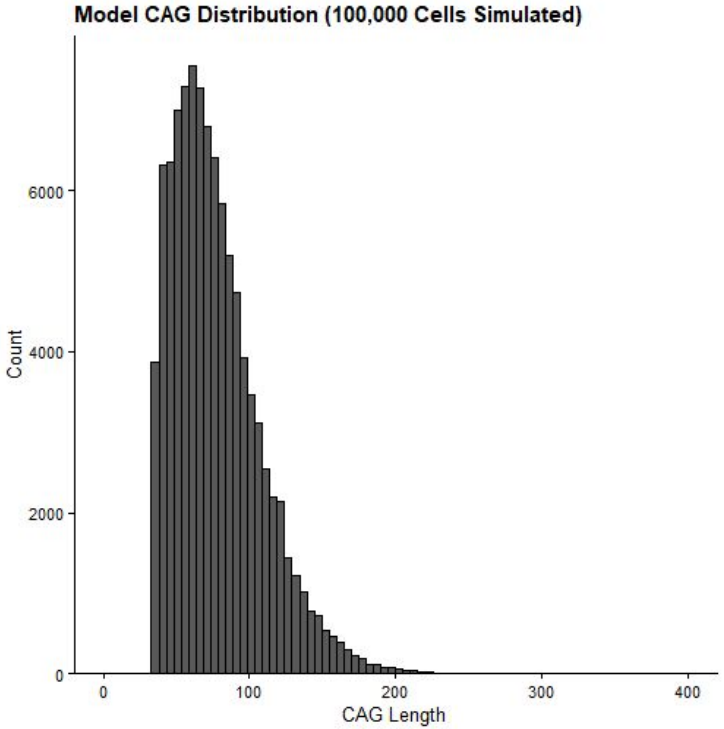
- Model one cell's CAG length at a time. Assume mutations are a random stochastic process with a certain rate that increases with longer CAG lengths.
- Each mutation has a probability of being an expansion and otherwise contraction, which increase/decrease the CAG length by 1.

We applied many variations that modeled the expansion rate as different functions of CAG length.

1 Higham, et al, HMG 2012



Simulate many cells



Likelihood Framework for Parameter Estimation

- To evaluate fit between the CAG length data and the distribution generated by the model, we used a likelihood-based framework.
- Let **CAG =** list of CAG lengths from the observed patient data.
- Given parameters **θ**, we run the model on a large number of cells (1,000,000), and then normalize the generated CAG length distribution. It will act as a probability density function **f**.
- We assume uniform priors for each parameter within reasonable ranges that we predetermine (essentially performing a grid search).

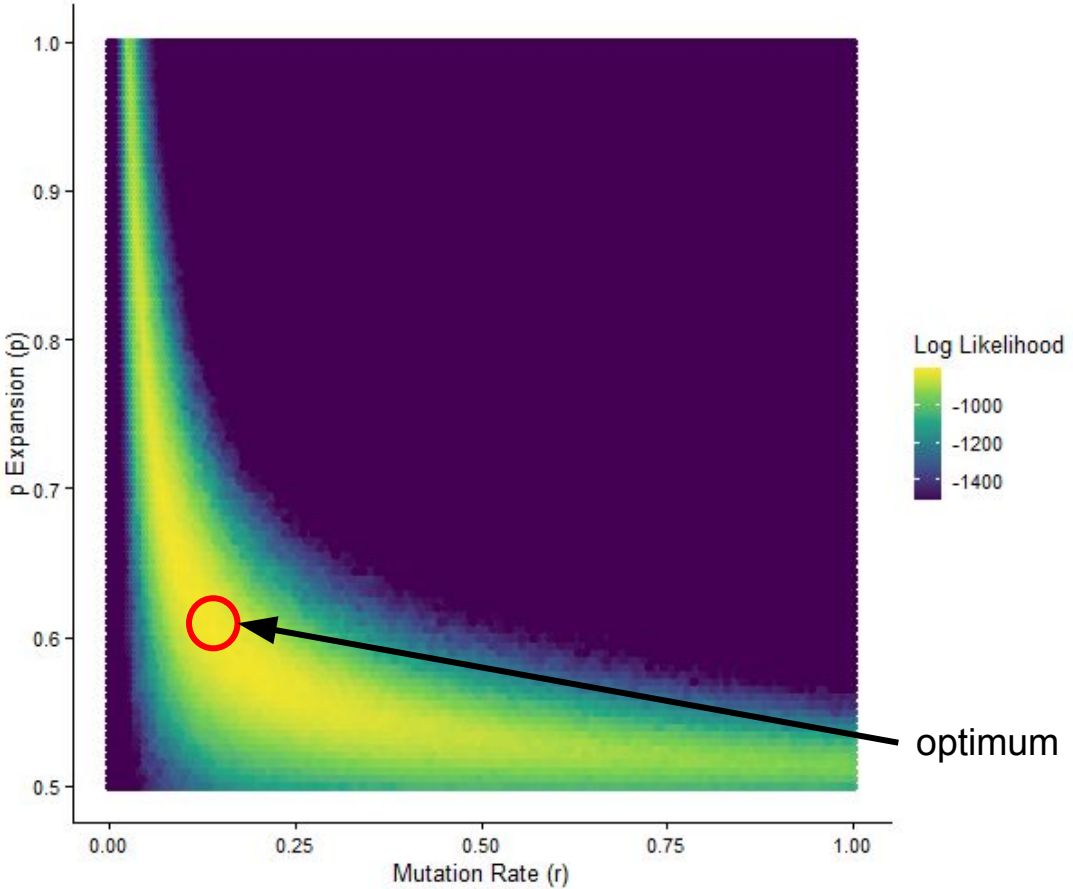
$$L(\theta|data) \sim P(data|\theta) = \prod_i f(CAG_i)$$

Likelihood (Cont.)

The plot to the right shows a grid search performed with a model consisting of two parameters.

Each point represents a model with specific values for its parameters, and colored by its likelihood.

The optimal set of parameters with highest likelihood is selected.

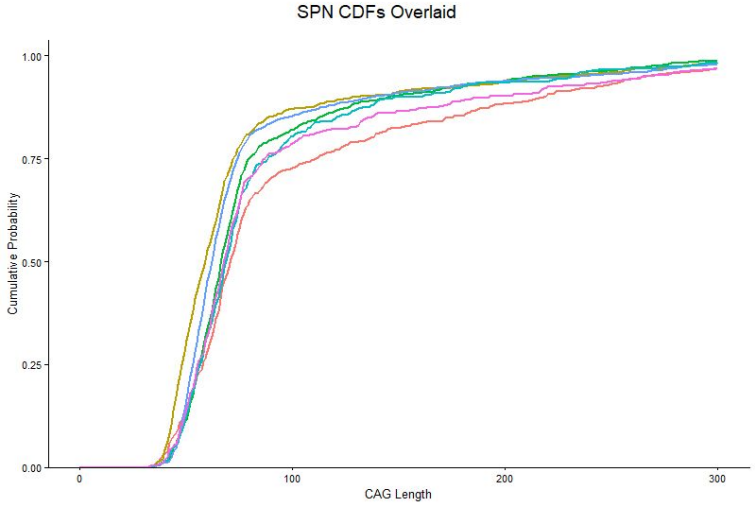


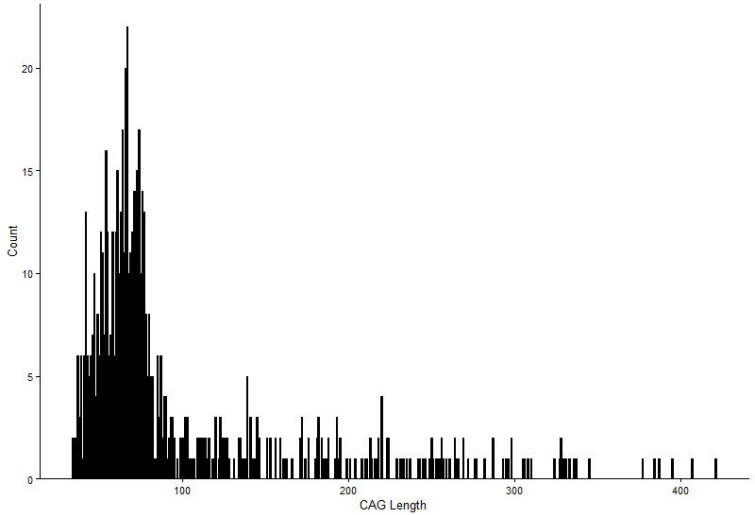
Two Phases of Expansion

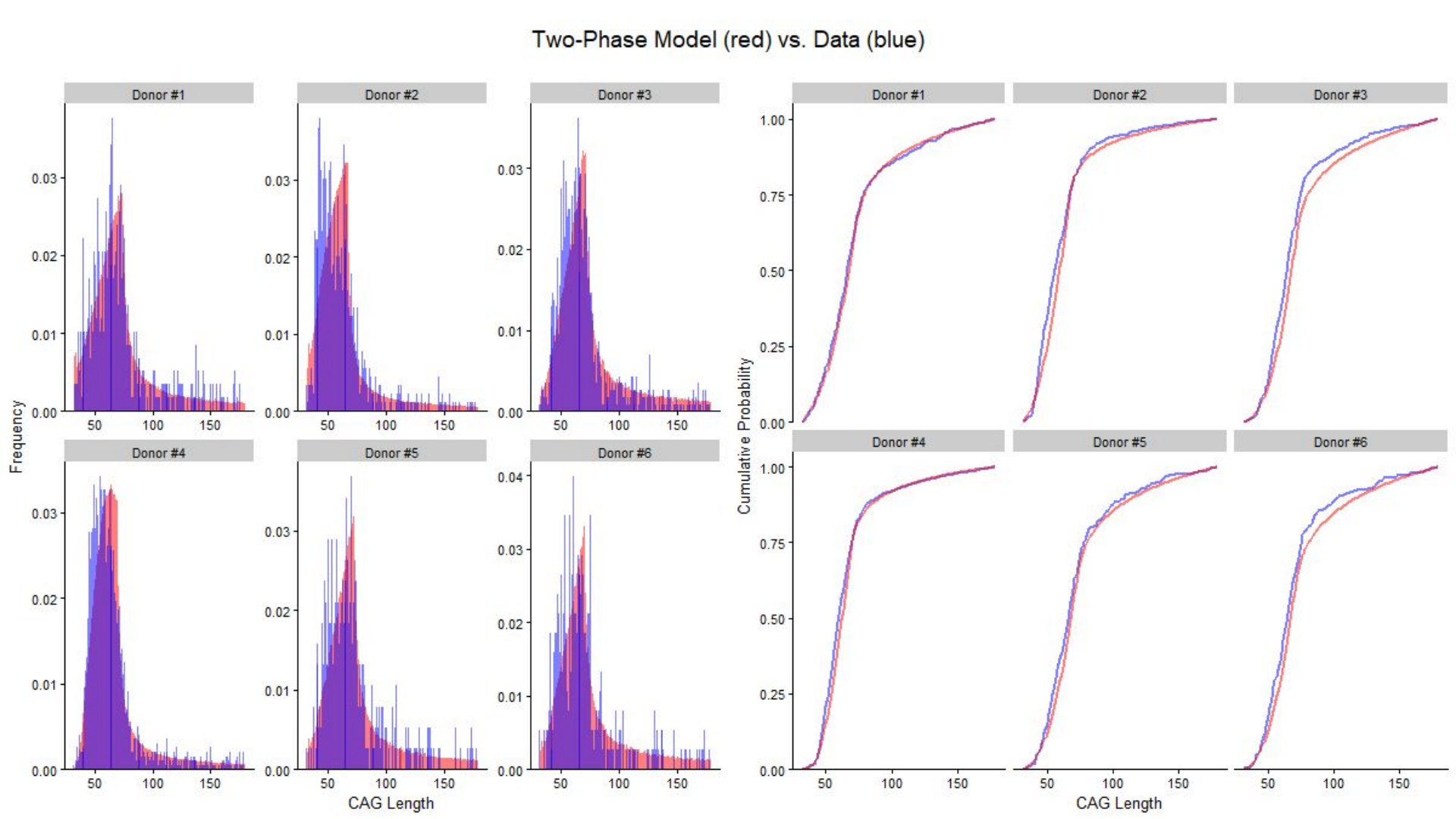
We applied a wide class of models, and found that models including two phases of expansion fit most optimally.

This can be seen in the data, where there is a sudden drop in frequency at around ~70 CAGs, and then a long tail.

We hypothesize that there could be a mechanism causing accelerated expansion once cells reach ~70 CAGs.

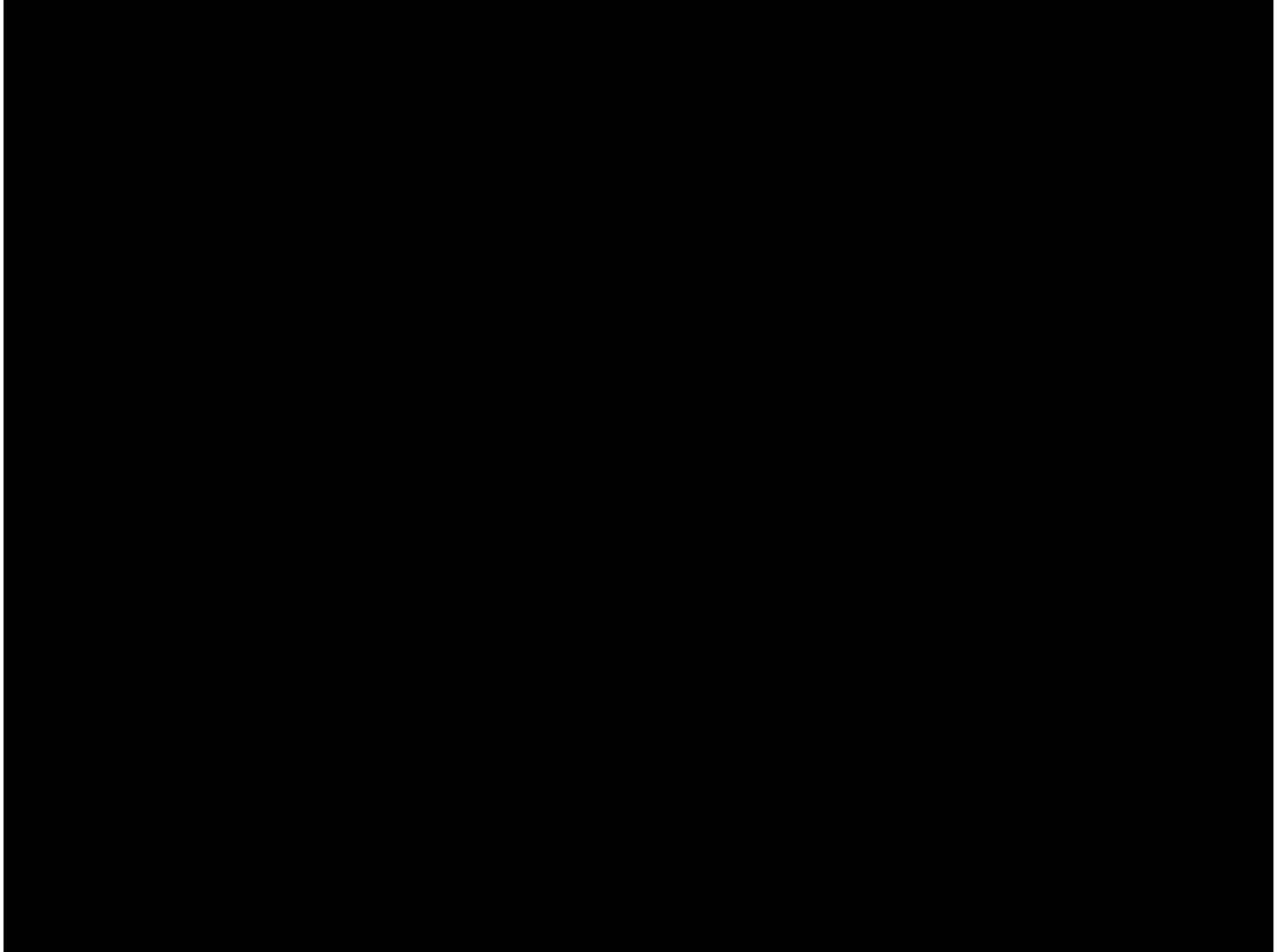




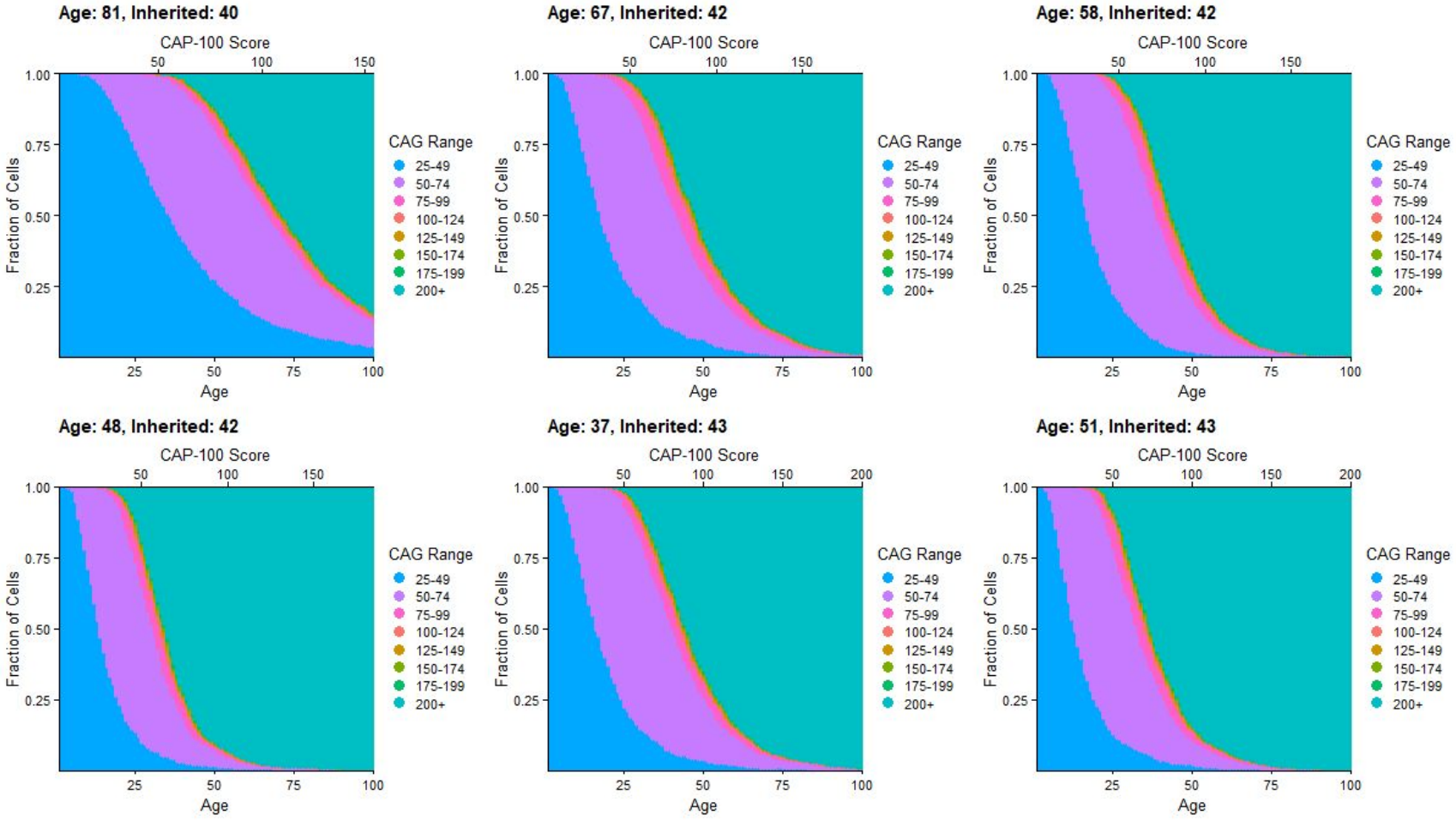

Two-Phase Model (red) vs. Data (blue)
Donor #1
Donor #2
Donor #3
Donor #4
Donor #5
Donor #6
Frequency
Cumulative Probability
CAG Length
CAG Length

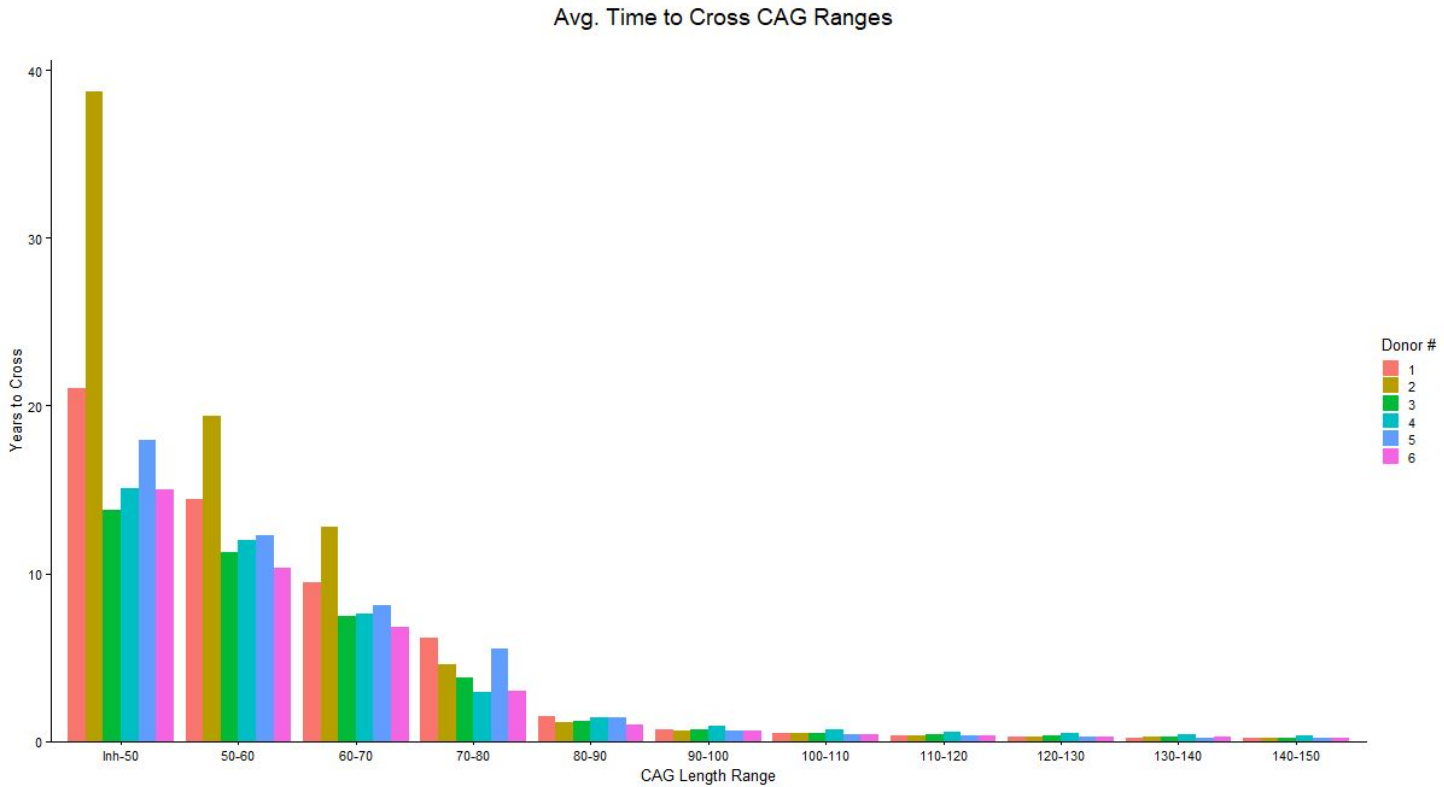
Applications of the Model

- The data collected only provide a single snapshot of the CAG lengths at the time of the patient's death.
- With the model, we can predict what the CAG lengths looked like at every stage of the patient's life.
- Specifically, we are interested in how long it takes cells to expand through certain CAG ranges, and how the rate of expansion changes.



Dynamics of Somatic Expansion





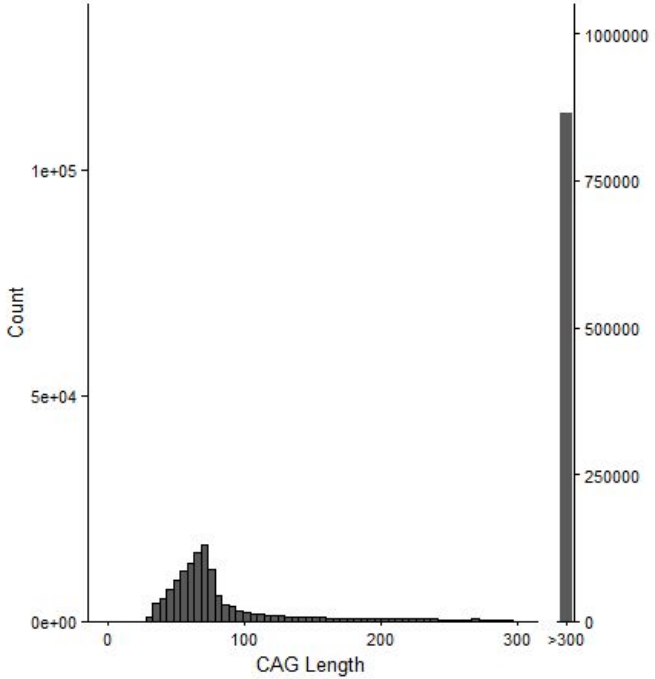
The plot above shows the average number of years it took each cell to cross certain CAG ranges based on the model.

Suggests that somatic expansion is a very slow process in lower CAG ranges but increases to extremely fast rates in the span of just a few years.

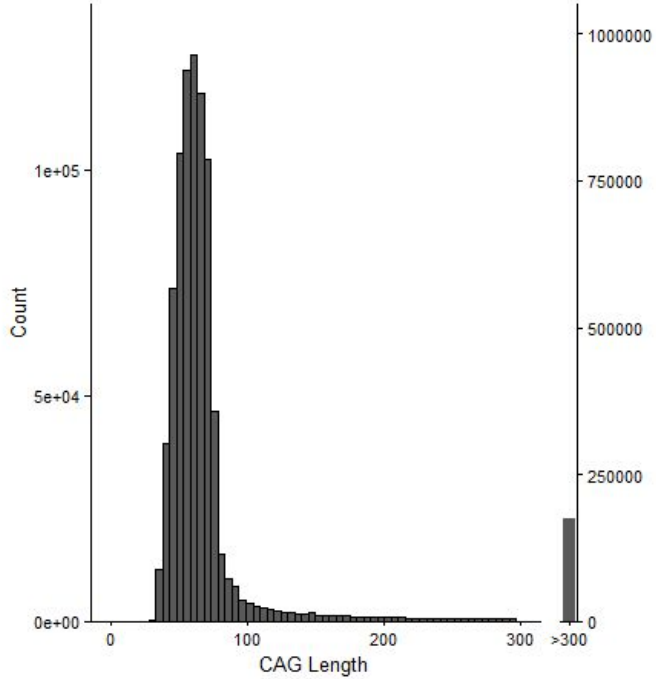
Implications of the Model

- The rate of CAG expansion increases as a superlinear function of CAG length.
    - The lab has found evidence that cells begin to exhibit transcriptional dysregulation at 180+ CAGs.
    - On average takes decades to go from inherited length to 70 CAGs, but only ~2 years to go from 180 CAGs to 500 CAGs.
    - There may only a small window of time in which cells exhibit “sick behavior.”
- Many therapeutic approaches are based on reducing *HTT* expression, but it may be that only a small fraction of cells are affected such toxicity at any time.
    - Rather, the model suggests therapeutics that attempt to slow somatic expansion may be much more effective.
    - A small reduction in expansion rate can have large effects in preventing cells from reaching the phase of accelerated expansion.

Normal Rate



Half Rate



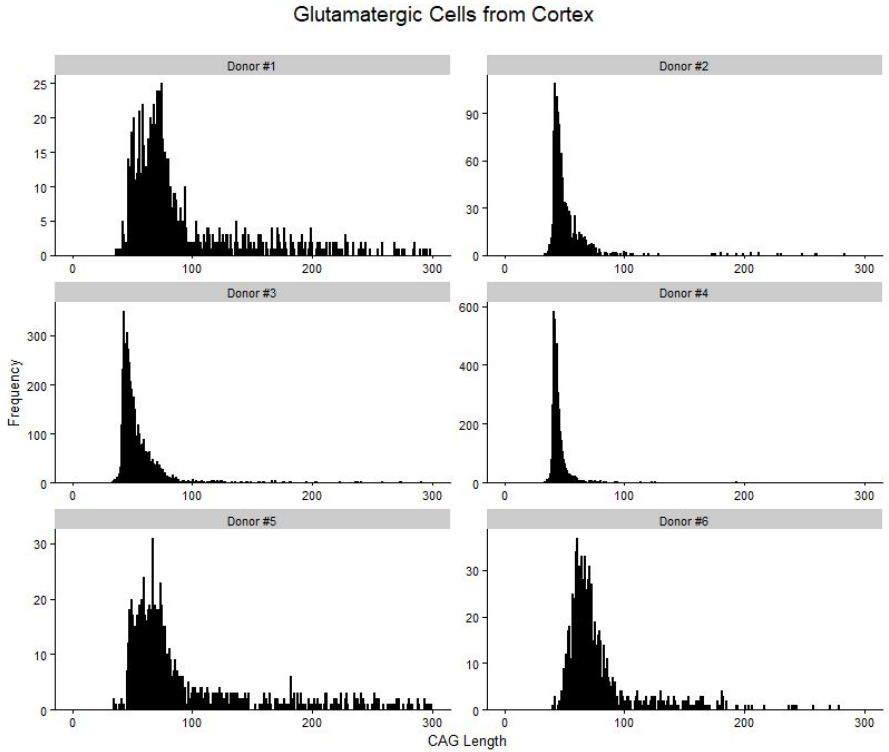
A model fit to one of the donors was re-run with the expansion rate halved, resulting in a massive reduction in overall expansion.

Potential Application in Clinical Trials

- Currently, Pfizer is working on a drug that targets somatic expansion.
- Clinical trials for testing such drugs can be time consuming and expensive.
- The model can help us design trials in a way that maximizes results in the smallest time window.
    - Suggests that we should give the therapeutic in patients in later stages of disease progression when the most cells are in the fast expansion phase.
    - Also informs how long we should expect to see results and the frequency to check on subjects.

Ongoing and Future Work

- Modeling cortical data has shown similar trends as SPNs so far.
- We plan on looking at a larger variety of cell types as well as cell subtypes.
    - Is the two-phase model is consistent across all cell types?
- We are also considering moving the model into a mathematical framework based on Markov Chains.
    - Would produce smoother theoretical distributions rather than empirical ones.
    - Computation speed up.



Acknowledgements

Bob Handsaker

Seva Kashin

Steven McCarroll

The McCarroll Lab

John Warner (from CHDI)

CHDI FOUNDATION