# Explainable Genetic Inheritance Pattern Prediction

Edmond Cunningham; Dana Schlegel; Andrew DeOrio
University of Michigan

## Introduction

Genetic specialists use high level pedigree features to identify the inheritance pattern of a disease.

We modeled pedigrees using latent state space models and inferred patient genotypes using relatives' phenotypes.

Our method allows for explainable predictions and the ability to add evaluate the effect of hypothetical evidence

## Mendelian Inheritance Patterns

**Autosomal Dominant**:
- Males and Females: AA, Aa, aa

**Autosomal Recessive**:
- Males and Females: aa, Aa, AA

**X-Linked Recessive**:
- Females: $X^AX^A$, $X^AX^a$, $X^aX^a$
- Males: $X^AY$, $X^aY$
- Unknown Sex: $X^AX^A$, $X^AX^a$, $X^aX^a$, $X^AY$, $X^aY$

Punnett squares were used to derive the model priors

| | | AA | | | Aa | | | aA | | aa | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | A | A | a | a | A | a | a | | |
| AA | A | AA | AA | AA | Aa | Aa | AA | Aa | Aa | | |
| | A | AA | AA | AA | Aa | Aa | AA | Aa | Aa | | |
| Aa | A | AA | AA | AA | Aa | Aa | AA | Aa | Aa | | |
| | a | aA | aA | aA | aa | aa | aA | aa | aa | | |
| aA | a | aA | aA | aA | aa | aa | aA | aa | aa | | |
| | A | AA | AA | AA | Aa | Aa | AA | Aa | Aa | | |
| aa | a | aA | aA | aA | aa | aa | aA | aa | aa | | |
| | a | aA | aA | aA | aa | aa | aA | aa | aa | | |

$$\left[ \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 0.5 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.5 & 0.5 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \end{bmatrix} \right]^T$$

## Genetic Inheritance as State Space Model

We can mapped genetic terms to state space terms:

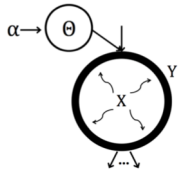
- Genotypes ➔ Latent States
- Phenotypes ➔ Observation States
- Mendel's Laws ➔ Prior Distribution
- Blood Test Result ➔ Latent State Evidence
- Rare Disease ➔ Roots Likely Not Carriers
- Human Intuition ➔ Hypothetical Evidence

We compared inheritance patterns using the marginal probability of the data:

$$P(Y; \{\text{AD}, \text{AR}, \text{XL}\}) = E_{\theta \sim P(\theta; \{\text{AD},\text{AR},\text{XL}\})}[P(Y|\theta)]$$

The full probabilistic model is:

$$\pi_0 \sim \text{Dirichlet}(\alpha^{\text{root}}),$$
$$\pi_{ij} \sim \text{Dirichlet}(\alpha_{ij}^{\text{transition}}),$$
$$L_i \sim \text{Dirichlet}(\alpha_i^{\text{emission}})$$
$$x^c \sim \text{Categorical}(\pi_{x^m x^f}),$$
$$y^c \sim \text{Categorical}(L_{x^c})$$




## Explainability

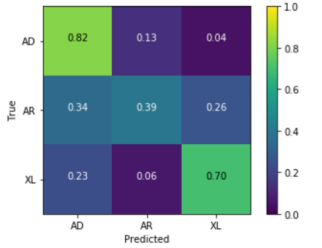
With each prediction, we can examine
- Inferred latent states
- Probability over states
- Latent features
  - *De novo* mutations
  - Incomplete penetrance instances
- Hypothetical situations
  - What if this person actually had this gene allele?

## Results

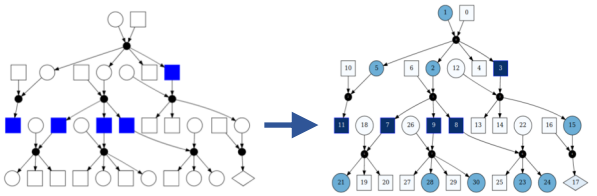
Our data was very noisy and had mislabeled data.

We still achieved the same accuracy in AD and XL as Schlegel et al.(2017).

AR was hindered by the abundance of pedigrees with only one labeled node.




We performed inference over an X-Linked pedigree to infer which family members are carriers



The shading on the right pedigree corresponds to the likelihood that a person is a carrier. The inferred latent states align with human intuition and as a result, the prediction can be trusted.

## Conclusions

**What can genetic specialists get from using our method?**
- Exact method of computing inheritance pattern probabilities
- Explainable predictions
- Ability to evaluate hypotheses over what reality