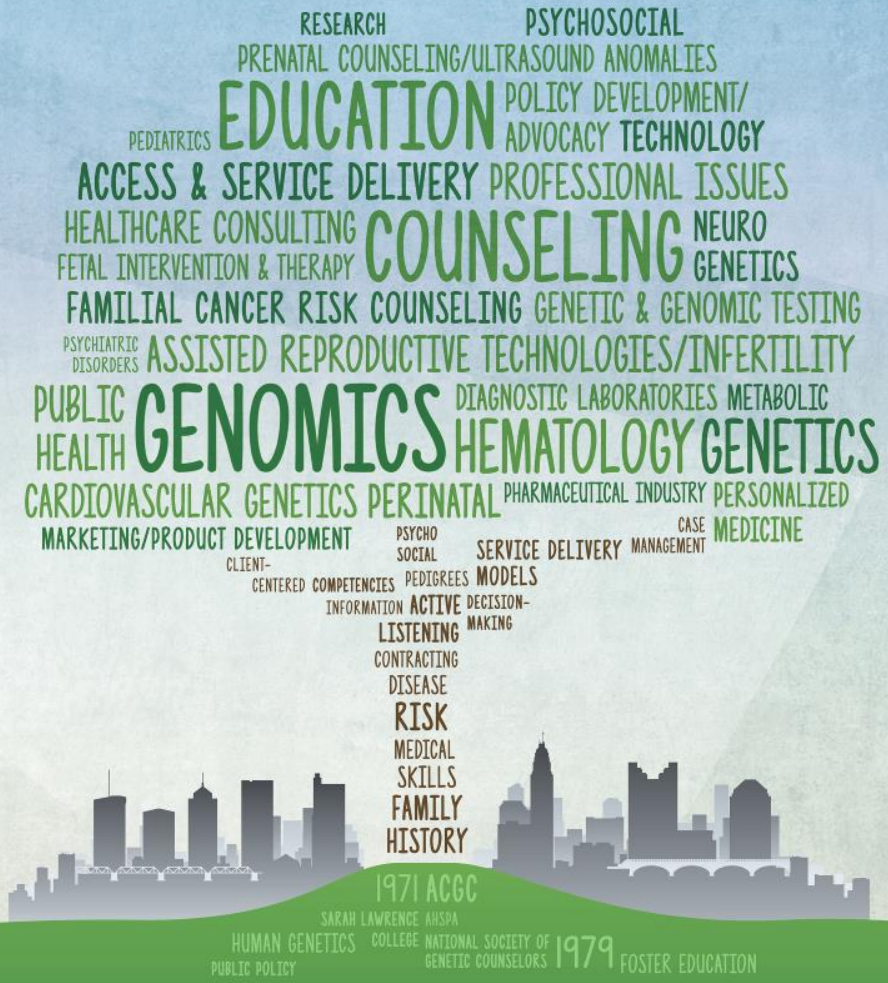
National Society of Genetic Counselors

36th Annual Conference

September 13-16, 2017

Greater Columbus Convention Center
Columbus, OH



• GROWING OUR PROFESSION TO NEW HEIGHTS •

COI Disclosure

No COI to disclose

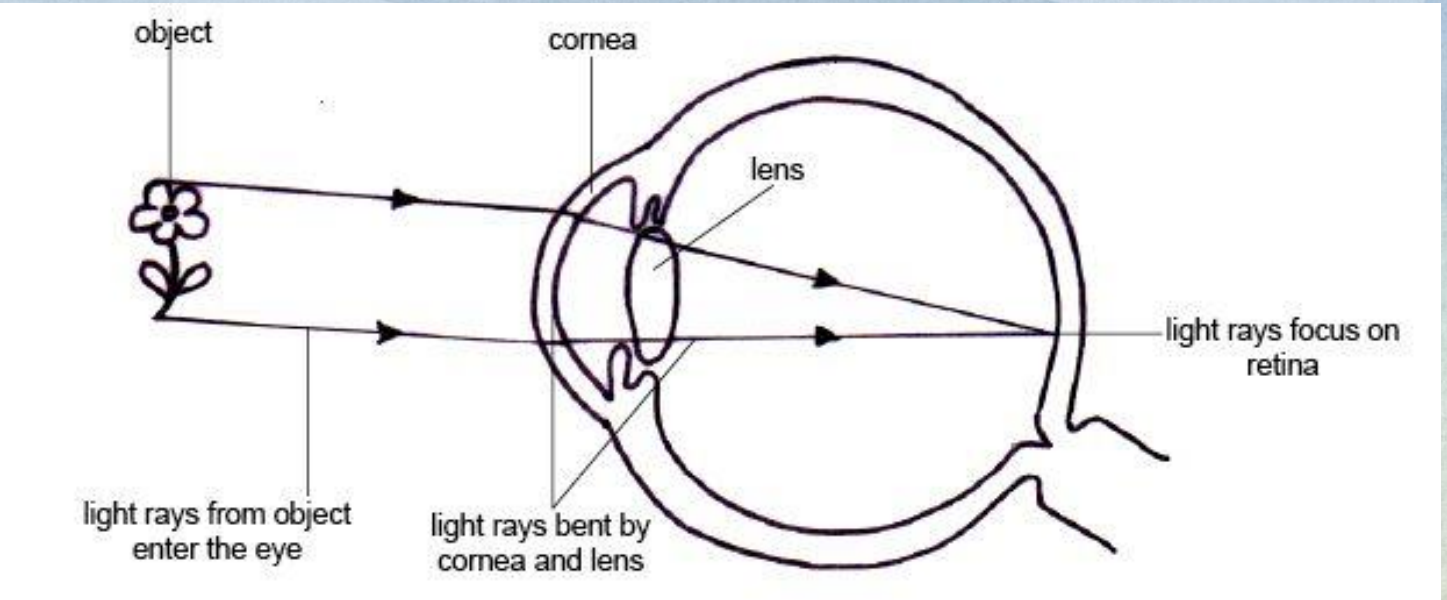
National Society of Genetic Counselors

36th Annual Conference
September 13-16, 2017
Greater Columbus Convention Center
Columbus, OH

Inheritance Pattern Prediction: An Ophthalmic Model for Digital Pedigree Feature Extraction and Machine Learning

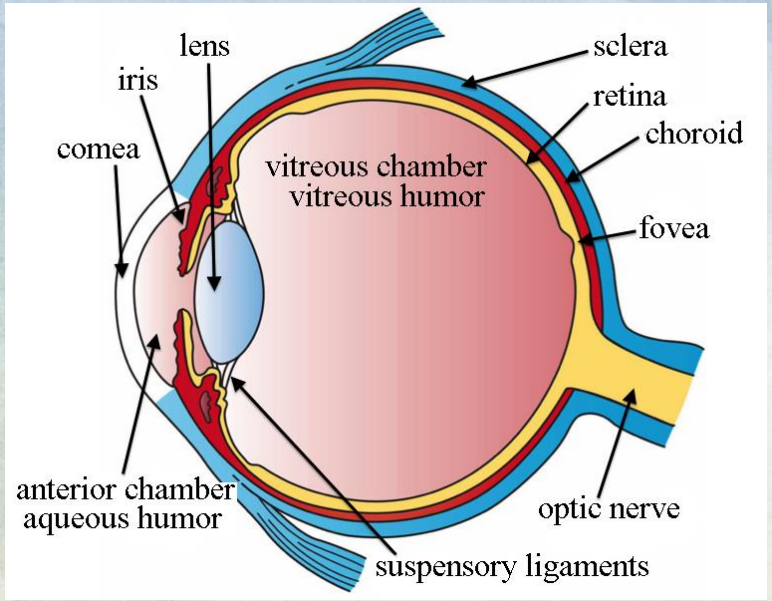
Dana Schlegel, MS, MPH, CGC; Edmond Cunningham; Xinghai Zhang; Yaman Abdulhak; Andrew DeOrio, PhD; K. Thiran Jayasundera, MD

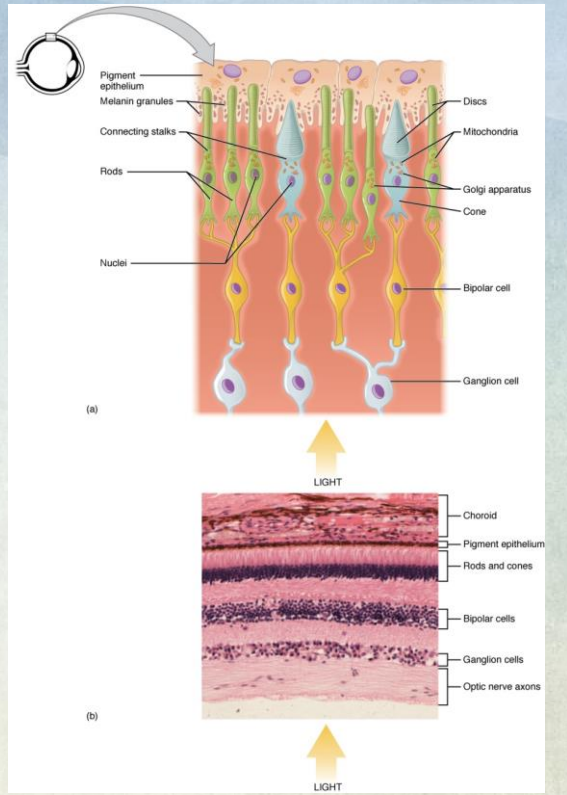
The eye: a brief overview



National Society of Genetic Counselors

Slightly more detail





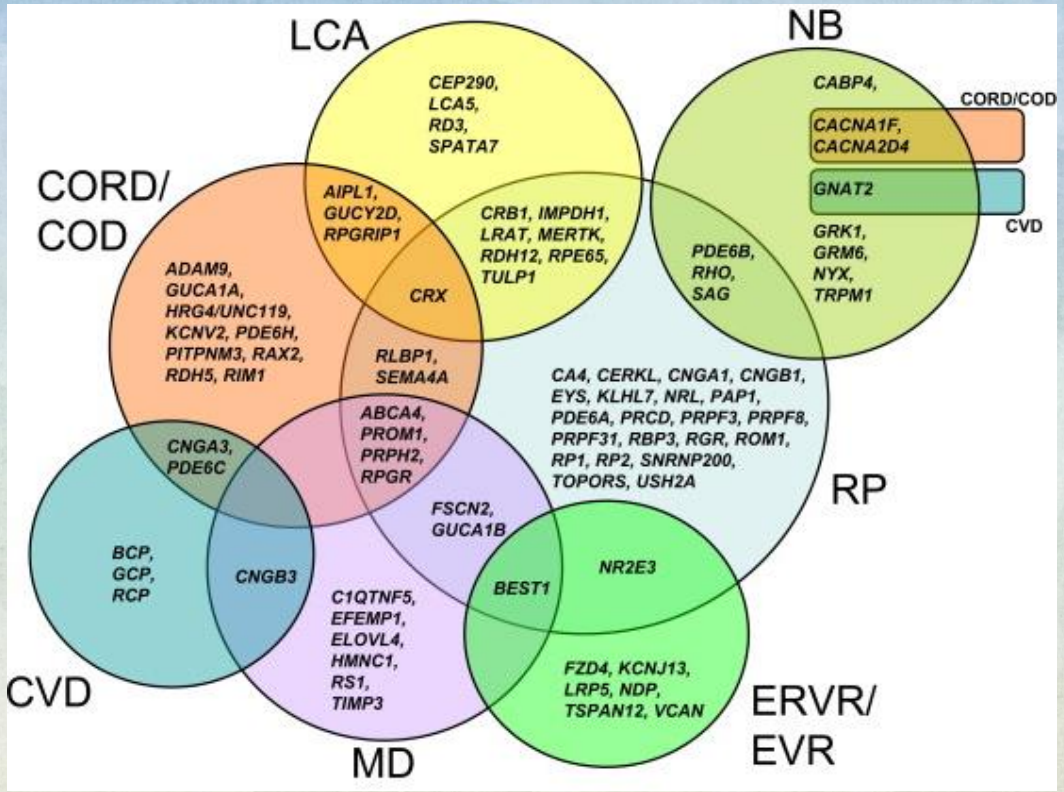
National Society of Genetic Counselors

Retinal dystrophies

- Inherited retinal degenerative diseases
  - Due to reduced or deteriorating function of cells of retina (ex. photoreceptors, retinal pigment epithelium)
  - Usually progressive, sometimes stationary
- Wide range of conditions
  - Retinitis Pigmentosa, Stargardt, Cone-rod dystrophy, Cone dystrophy, Choroideremia, Leber Congenital Amaurosis, Usher, Bardet-Biedl syndrome…
- Genetically complicated/diverse
  - Clinical heterogeneity, genetic heterogeneity, variable expressivity, incomplete penetrance, some genes with multiple patterns of inheritance

National Society of Genetic Counselors

Retinal Dystrophies

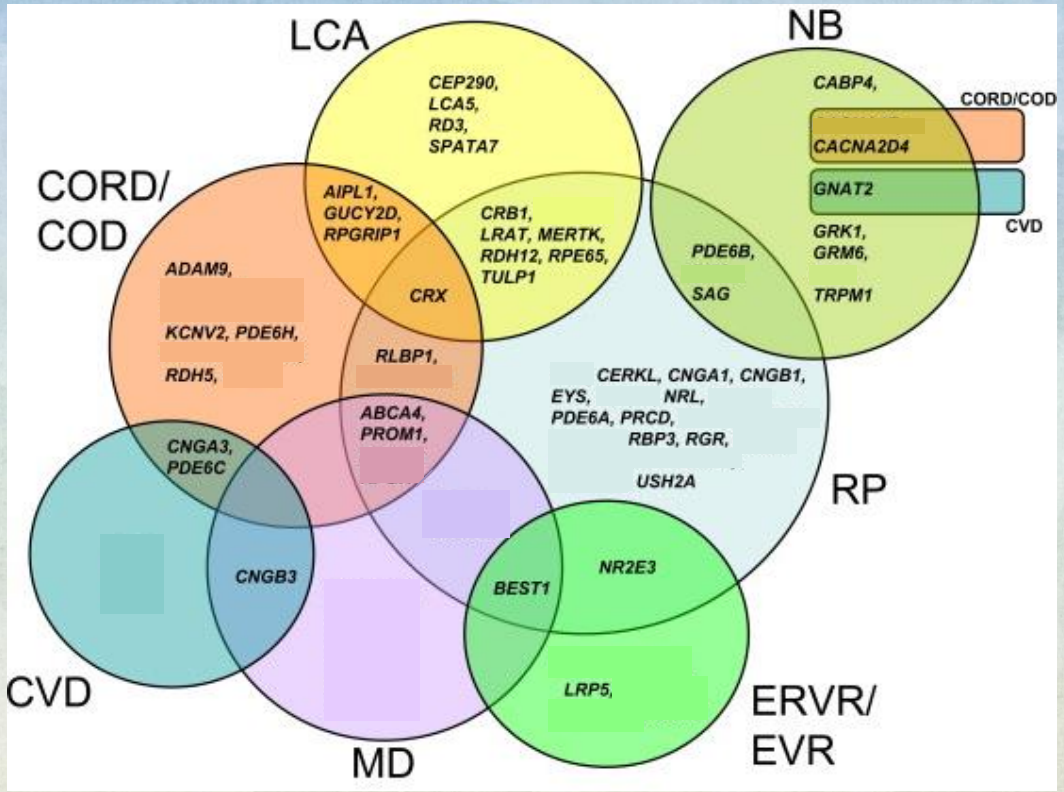


Berger W et al, 2010.

National Society of Genetic Counselors

Retinal Dystrophies

Autosomal Recessive (AR)

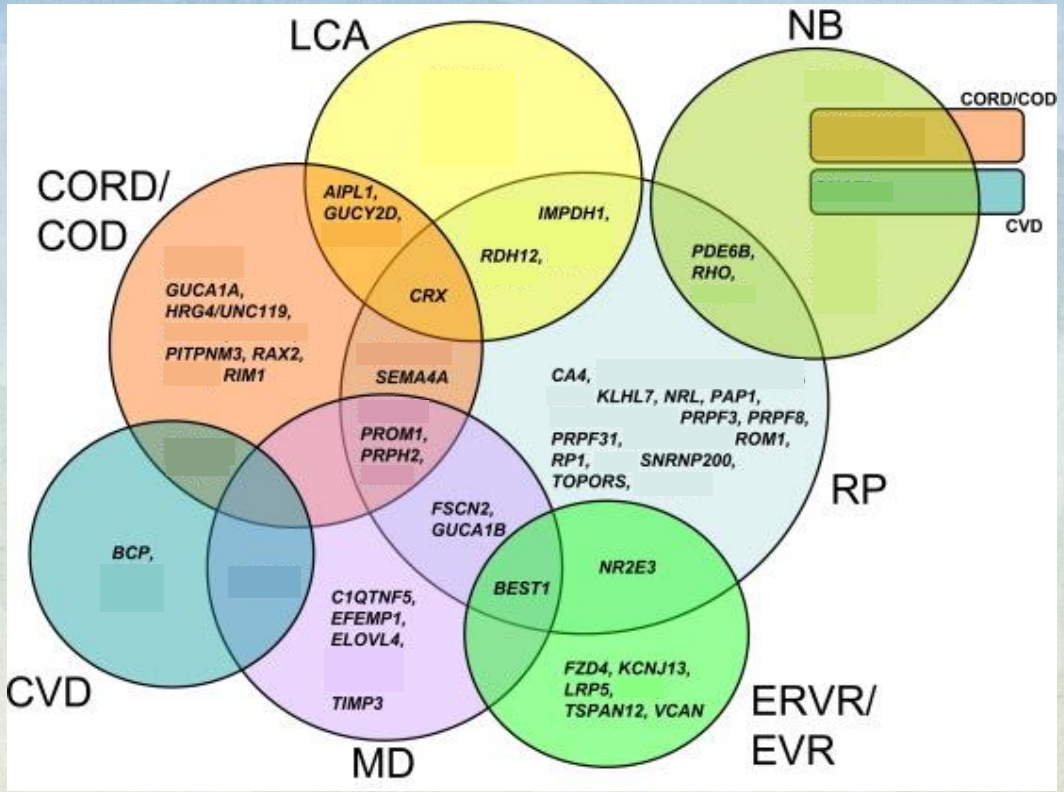


Adapted from Berger W et al, 2010.

National Society of Genetic Counselors

Retinal Dystrophies

Autosomal Dominant (AD)

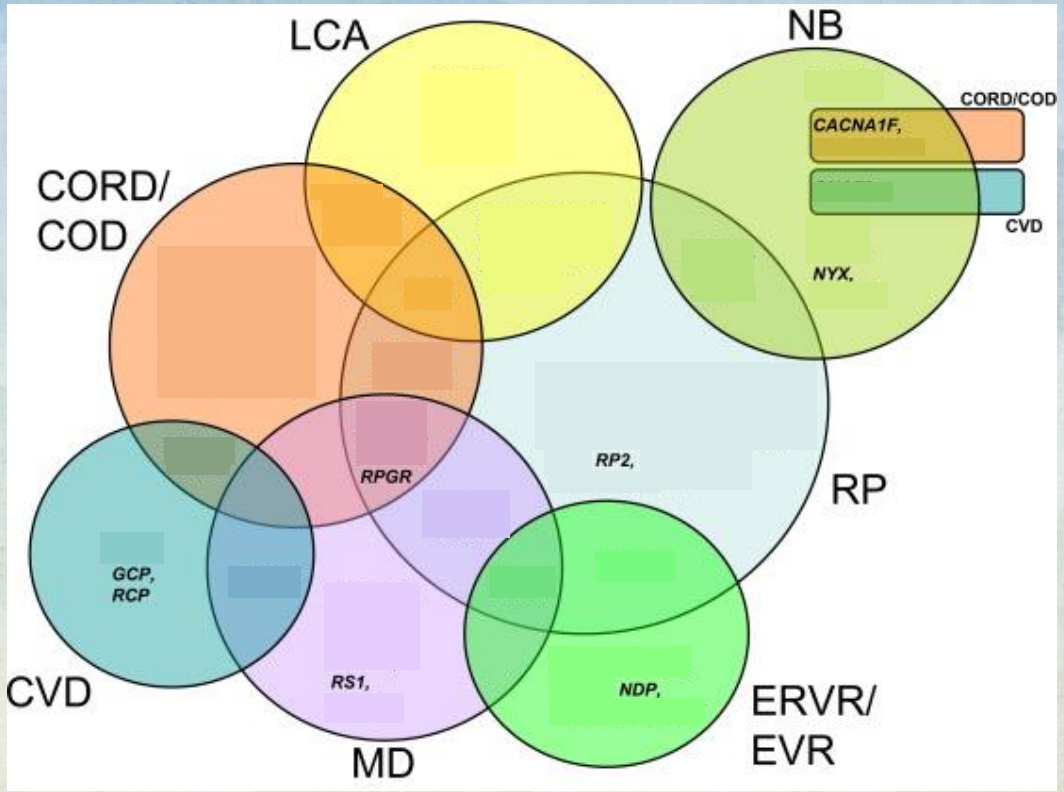


Adapted from Berger W et al, 2010.

National Society of Genetic Counselors

Retinal Dystrophies

X-linked (XL)



Adapted from Berger W et al, 2010.

National Society of Genetic Counselors

Inheritance Pattern Prediction

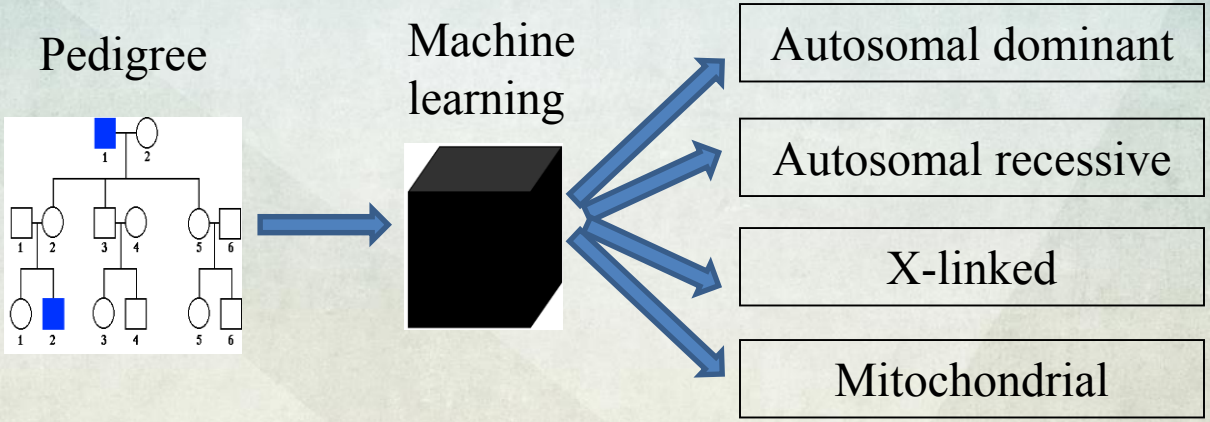
- May inform likely diagnosis
- Can guide appropriate genetic testing
- Allows calculation of likely risks to relatives
- Required component of data collection for some retinal dystrophy studies

As far as we are aware, there is no current algorithm to predict pattern of inheritance for a given patient, and not all retinal dystrophy clinics have genetics services

National Society of Genetic Counselors

Aim

- Create a machine learning algorithm whose input is patient family history information and whose output is likely pattern of inheritance
- Used retrospective chart review on patients with genetically-proven retinal dystrophies



National Society of Genetic Counselors

Data collection

- Kellogg Eye Center retinal dystrophy patients with genetic diagnosis
- Family history obtained by genetic counselors (and, in rare cases, retinal dystrophy specialists) as a part of routine patient care
- Information collected by engineering and medical students trained by genetic counselors and retinal dystrophy specialists
- Pedigrees converted into digital computer-readable form

National Society of Genetic Counselors

Data collection methodology

- Students trained in predicting pattern of inheritance based on interpretation of pedigree appearance evaluated likely pattern of inheritance for each patient (277 patients)

- Answers to 12 questions about family history were collected from each patient's pedigree and analyzed with machine learning (100 patients)

- Answers to the same 12 questions were collected through computer feature extraction of a digitized pedigree and analyzed with machine learning (90 patients)
  - Included tolerance for user input error

(Overlap of 70 patients between the three cohorts)

National Society of Genetic Counselors

Family history features

	Question	Possible Answers
1	Is more than one generation affected?	Yes/No
2	Do any affected males have affected sons?	Yes/No
3	Do any affected males have affected daughters?	Yes/No
4	Are there any unaffected individuals who are "skipped"? (Their parents or siblings or grandparents are affected and children or grandchildren are affected, but they themselves are unaffected.)	1. No 2. Yes - females only are skipped 3. Yes - at least some males are skipped
5	Are any siblings of the patient affected?	1. No 2. Yes, and no other relatives are affected 3. Yes, and other relatives are also affected
6	Are any cousins of the patient affected?	1. No 2. Yes - maternal cousins only 3. Yes - paternal cousins only 4. Yes - maternal and paternal cousins
7	Are both males and females affected?	1. Yes 2. No - only males 3. No - only females
8	Is onset of disease < or = 20yrs in males?	Yes/No
9	Do any females have asymmetric disease?	Yes/No
10	In general, do females have less severe or later onset of disease?	Yes/No
11	Is there more than one retinal diagnosis in the family? (ex. Stargardt and Pattern Dystrophy)	Yes/No
12	Is consanguinity present?	Yes/No

National Society of Genetic Counselors

Family history features

	Question	Possible Answers
1	Is more than one generation affected?	Yes/No
2	Do any affected males have affected sons?	Yes/No
3	Do any affected males have affected daughters?	Yes/No
4	Are there any unaffected individuals who are "skipped"? (Their parents or siblings or grandparents are affected and children or grandchildren are affected, but they themselves are unaffected.)	1. No 2. Yes - females only are skipped 3. Yes - at least some males are skipped
5	Are any siblings of the patient affected?	1. No 2. Yes, and no other relatives are affected 3. Yes, and other relatives are also affected
6	Are any cousins of the patient affected?	1. No 2. Yes - maternal cousins only 3. Yes - paternal cousins only 4. Yes - maternal and paternal cousins
7	Are both males and females affected?	1. Yes 2. No - only males 3. No - only females
8	Is onset of disease < or = 20yrs in males?	Yes/No
9	Do any females have asymmetric disease?	Yes/No
10	In general, do females have less severe or later onset of disease?	Yes/No
11	Is there more than one retinal diagnosis in the family? (ex. Stargardt and Pattern Dystrophy)	Yes/No
12	Is consanguinity present?	Yes/No

National Society of Genetic Counselors

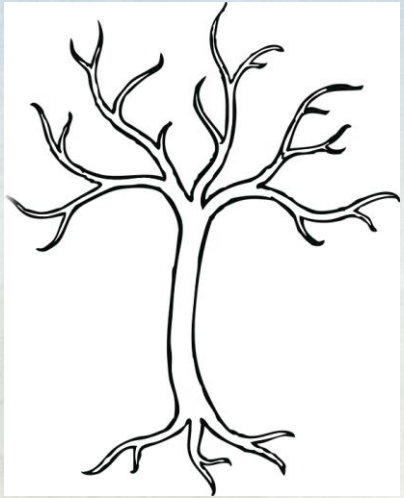
Family history features

	Question	Possible Answers
1	Is more than one generation affected?	Yes/No
2	Do any affected males have affected sons?	Yes/No
3	Do any affected males have affected daughters?	Yes/No
4	Are there any unaffected individuals who are "skipped"? (Their parents or siblings or grandparents are affected and children or grandchildren are affected, but they themselves are unaffected.)	1. No 2. Yes - females only are skipped 3. Yes - at least some males are skipped
5	Are any siblings of the patient affected?	1. No 2. Yes, and no other relatives are affected 3. Yes, and other relatives are also affected
6	Are any cousins of the patient affected?	1. No 2. Yes - maternal cousins only 3. Yes - paternal cousins only 4. Yes - maternal and paternal cousins
7	Are both males and females affected?	1. Yes 2. No - only males 3. No - only females
8	Is onset of disease < or = 20yrs in males?	Yes/No
9	Do any females have asymmetric disease?	Yes/No
10	In general, do females have less severe or later onset of disease?	Yes/No
11	Is there more than one retinal diagnosis in the family? (ex. Stargardt and Pattern Dystrophy)	Yes/No
12	Is consanguinity present?	Yes/No

National Society of Genetic Counselors

Machine learning methodology

Gradient-Boosted Tree



Machine learns appropriate weight for each branch

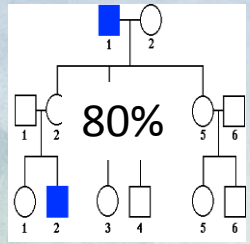
Decision tree

National Society of Genetic Counselors

Machine learning methodology

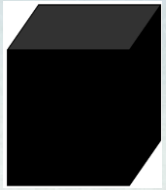
80/20 training/testing split

Training

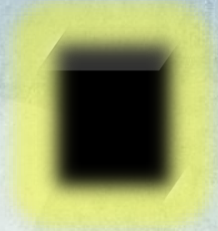


Inheritance pattern

Machine learning



Classifier

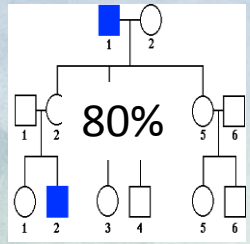


National Society of Genetic Counselors

Machine learning methodology

80/20 training/testing split

Training

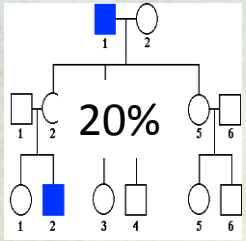


Inheritance pattern

Machine learning

Classifier

Testing



Predicted pattern of inheritance

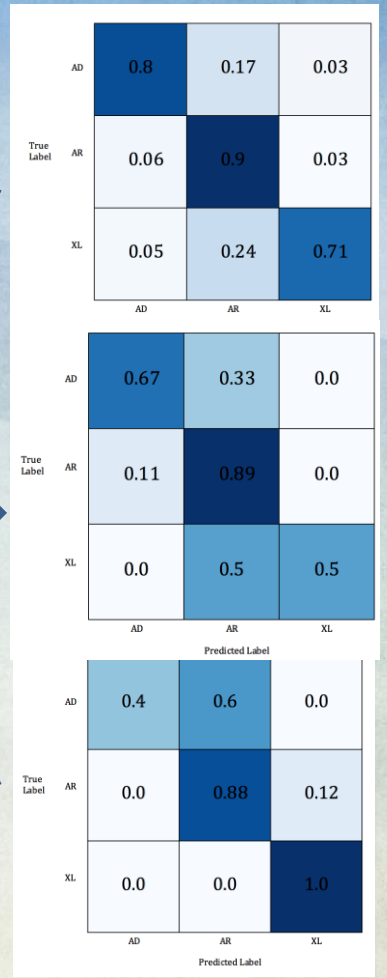
National Society of Genetic Counselors

Results

Method	Accuracy	Standard Deviation
Human-predicted	84%	--
Machine learning with human-entered answers	78%	7.5%
Machine learning with computer-extracted answers	76%	9.8%

National Society of Genetic Counselors

Method	Accuracy	Standard Deviation
Human-predicted	84%	--
Machine learning with human-entered answers	78%	7.5%
Machine learning with computer-extracted answers	76%	9.8%



National Society of Genetic Counselors

Challenges

- Small dataset
  - Limited to patients with definitive genetic diagnosis
- Machine learning, but human-written questions
  - Our assumptions about the most important questions to ask may not always be correct
  - Is it better to ask more questions or fewer?
- Machines can make mistakes, too
  - Attributing importance to unimportant features (worse with small dataset)
- Perfect prediction is impossible
  - Ex. Isolated cases

National Society of Genetic Counselors

Future Directions

- Collect more data from other institutions
  - Machine learning relies on large datasets for sufficient training
- As data collection increases, adjust questions that are informative/non-informative
  - Our expectations about what questions would be most useful might not have been correct
- Use machine learning directly on pedigree, without answering questions
  - Use statistical analysis (Bayesian inference, hidden Markov models) to supplement or substitute for machine learning methodology

National Society of Genetic Counselors

- University of Michigan Kellogg Eye Center
  - Thiran Jayasundera, MD
  - Kari Branham, MS, CGC
  - Naheed Khan, PhD
  - Abigail Fahim, MD, PhD
  - John Heckenlively, MD
  - Eman Al-Sharif
- eyeGENE research project
- University of Michigan Computer Science & Engineering Department
  - Andrew DeOrio, PhD
  - **Edmond Cunningham**
  - Xinghai Zhang
  - Yaman Abdulhak
- Funding
  - University of Michigan Multidisciplinary Program (MDP)
  - Jayasundera startup grant

National Society of Genetic Counselors