RetDegenDx: A retinal dystrophy genetic diagnosis prediction tool

KELLOGG EYE CENTER
UNIVERSITY OF MICHIGAN HEALTH SYSTEM

Schlegel, Dana[1], Benjamin Katz[2], Bichucher, Vittorio[2], Starbuck, Richmond[2], Xu, Wei[2], Durrah, Jacob[2], Al-Sharif, Eman[1], DeOrio, Andrew[3], Khan, Naheed W.[1], Jayasundera, Kanishka T.[1]

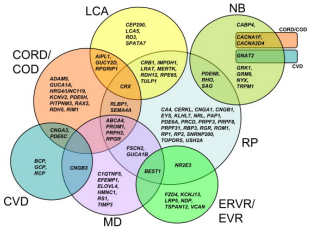
[1]Ophthalmology and Visual Sciences, [2]Multidisciplinary Design Program, and [3]Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI

BACKGROUND

Retinal dystrophies are a group of conditions that affect 1/2000 individuals.[1] They are diagnosed based on the reported symptoms, the results of clinical testing, the course of disease, and the predicted mode of inheritance.[2] There are currently only a handful of specialists in the United States who can effectively diagnose and manage patients with retinal dystrophies.

These conditions are clinically and genetically heterogeneous. (Figure 1) Over 170 genes and thousands of different mutations have been implicated as disease-causing in individuals with various retinal dystrophies.[3] Identifying the genetic diagnosis for a given patient is important for a number of reasons. It can confirm the diagnosis, direct clinical management, provide a more accurate prognosis, inform genetic counseling, and identify patients for whom molecular-based therapy might be available.[2]

Genetic testing, however, is labor-intensive and expensive and is therefore only performed on a subset of patients.[2] The cost of testing depends which and how many genes are analyzed and can be anywhere from several hundred to several thousand dollars. Thus, there is a need for improved diagnostic tools to help determine a patient's primary clinical diagnosis and the appropriate targets for genetic testing. In addition, even when genetic testing is affordable, results may be inconclusive, and new methods are necessary to help interpret these test reports and identify which gene variants are most likely to be disease-causing.



Berger et al, 2010

Figure 1. Genetic causes of retinal dystrophies

PURPOSE

Scientists in other fields have begun to recognize the utility of computer based applications for diagnosing different diseases.[4,5] Through the creation of a large database of mutation-proven patients and the utilization of machine learning methodology, we are therefore creating a diagnostic tool that predicts the likely mutated gene responsible for the patient's diagnosis.

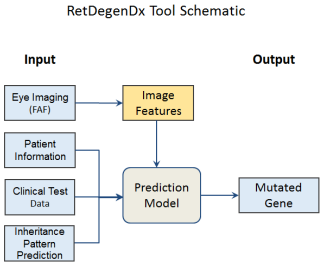


Figure 2. Schematic representation

Our purpose is to develop a machine learning program that predicts the most likely genetic cause of a patient's retinal dystrophy using input information about patient demographics, electroretinogram (ERG) response, visual field, pattern of inheritance, and fundus autofluorescence (FAF) features. This program may help to inform appropriate genetic test panels to order and can be used as a tool with which to interpret genetic testing results, along with existing resources such as PolyPhen, SIFT, and reported causative mutations in the literature.

METHODS

Data on patient age, sex, ERG response, visual field, family history, FAF imaging, and genetic diagnosis was collected on 152 patients seen at the Kellogg Eye Center. Alter filtering out mutated genes that affected fewer than 5 patients, 102 patients were usable for machine training purposes. Machine learning algorithms were developed to predict the genes most likely to be mutated and causing the observed clinical features for a given patient. Multiple algorithms were applied, and the support vector machine (SVM) with linear and radial basis function (RBF) kernel was shown to perform best. Machine learning used 80/20 training/testing splits of the data. Imputation techniques were applied to compare the classification performance with that of a baseline classifier.

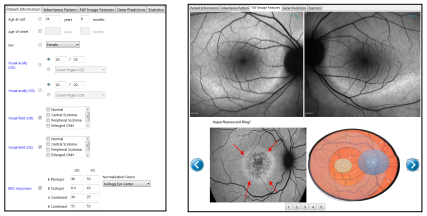


Figure 3. Input fields

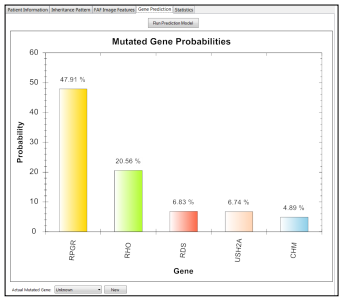


Figure 4. Output field

RESULTS

A prototype has been created that has greater than 60% accuracy for predicting the causative mutated gene in a given patient's sample.

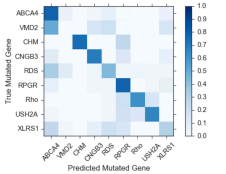


Figure 5. Accuracy by gene

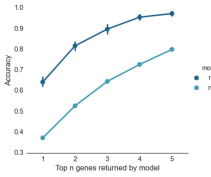


Figure 6. Accuracy of machine learning vs naïve model

Limitations

Data is currently limited, and many mutated genes are unique to fewer than 5 patients at the Kellogg Eye Center. This prototype is currently being shown to other retinal dystrophy clinics to obtain more patient data from other institutions. With the continued collaboration of outside institutions, the functionality of the algorithm will be improved.

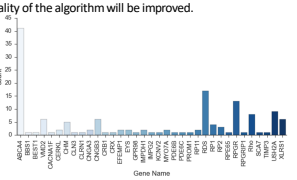


Figure 7. Current database – number of patients per causal mutated gene

CONCLUSIONS

The generation of this machine learning program provides physicians with a prediction of a causative mutated gene when evaluating a new patient in their clinic. This may help ophthalmologists refine their clinical diagnosis and identify the most relevant genetic testing to order. Additionally, genetic test results can often be difficult to interpret. Programs such as PolyPhen and SIFT help to determine whether or not a variant detected in a gene is truly pathogenic. The RetDegenDx program is one further tool that can be used to help predict the gene in which mutations are most likely to cause the clinical features of the patient. When used alongside existing programs, it can help to improve the interpretation of genetic test results.

	PolyPhen (Polymorphism Phenotyping)	SIFT (Sorting Intolerant From Tolerant)	RetDegenDx
Method	Protein structure/ function	Protein structure/ function	Clinical features, inheritance pattern
Sensitivity	68%[6]	69%[6]	68-72%

Figure 8. Accuracy of different methods for predicting the likelihood that a given gene is responsible for a patient's phenotype when a variant in that gene has been identified in the patient's sample

REFERENCES

1. Gilbert, C., Foster, A. (2011). Childhood blindness in the context of VISION 2020- The Right to Sight. *Bulletin of the World Health Organization*. 79(3):227-32.
2. Berger, W., Kloeckener-Gruissem, B., Neidhardt, J. (2010). The molecular basis of human retinal and vitreoretinal diseases. *Progress in Retinal and Eye Research* Sep;29(5):335-75.
3. O'Sullivan, J., Mullaney, B.G., Bhaskar, S.S., Dickerson, J.E., Hall, G., O'Grady, A., Webster, A., Ramsden, S.C., & Black, G.C. (2012). A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. *Journal of Medical Genetics*. May;49(5):322-6.
4. Karnowski, T.P., Giancardo, L., Aykac, D., Tobin, K. W., Li, Y., Garg, S., Abramoff, M. D., Tennant, M. T., & Chaum, E. Automating the Diagnosis, Stratification, and Management of Diabetic Retinopathy Using Content-Based Image Retrieval in an Ocular Telehealth Network. *Computational Analysis of the Human Eye with Applications* (pp. 417-453). River Edge, NJ: WSPC.
5. Avitzur, O. (2007). Blogs, Wikis, and Other Cutting Edge Technologies for Neurology. *Neurology Today*. Feb;7(3): 24–26.
6. Flanagan SE, Patch AM, Ellard S. (2010). Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. Genet Test Mol Biomarkers. Aug;14(4):533-7.