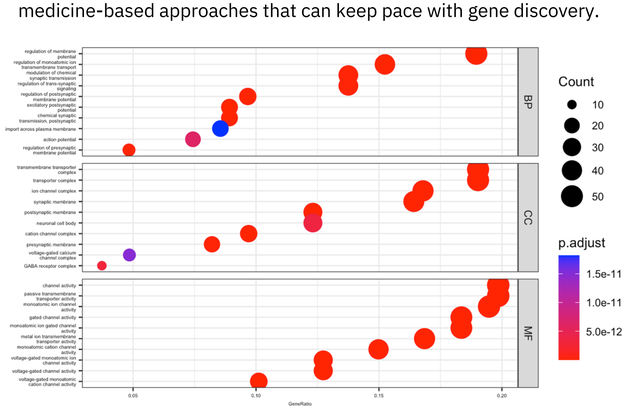
Ontology-Driven Approach to Facilitate Genetic Epilepsy Syndrome Diagnostic Workups

BIOMEDIN 210/CS 270 Class Project
Yi Li, Yuyi Zhang, Minh Vu, Ayush Singla

Background

Epilepsy is a common neurological disorder with a significant genetic component, yet the genetic etiology remains elusive in most cases. Identifying the underlying genetic syndrome of epilepsy is crucial for personalized treatment and improved patient outcomes. However, the rapid discovery of new epilepsy-related genes poses a challenge for physicians to keep pace.

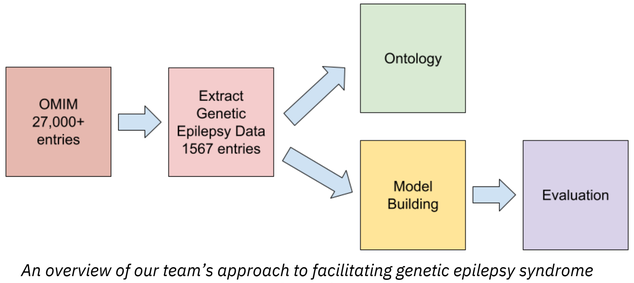
Our project aims to bridge this gap by developing an ontology-driven system that leverages the knowledge from the OMIM database and Bayesian modeling. The OMIM (Online Mendelian Inheritance in Man) database is an online catalog of human genes and genetic disorders. Using OMIM data, our system defines the probability of genetic epilepsy syndromes based on clinical symptoms, thus assisting physicians in formulating diagnostic hypotheses and proposing appropriate workups. By integrating an ontology framework with probabilistic reasoning, our approach enhances the understanding of complex relationships between symptoms and genetic syndromes and enables precision medicine-based approaches that can keep pace with gene discovery.



The GO enrichment analysis of genes identified in the genetic epilepsy syndromes indicated that they are involved in various domains of functions and pathways.

*__BP__ for Biological Process, __MF__ for Molecular Function, and __CC__ for Cellular Component.

Project Pipeline



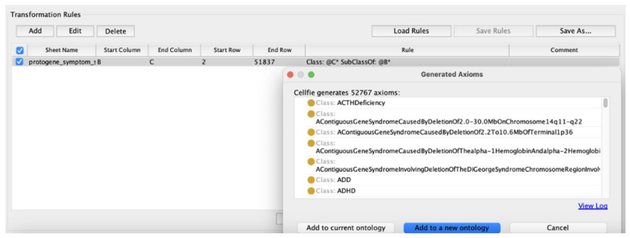
An overview of our team's approach to facilitating genetic epilepsy syndrome diagnosis.

Ontology

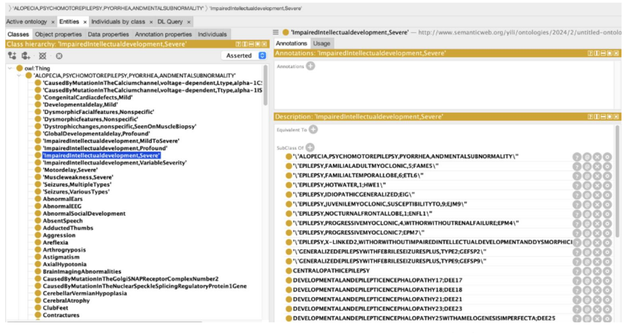
The diagram below shows how we used the processed genetic epilepsy data to construct a simple ontology.

	A	B	
51827	616230	EPILEPSY, PROGRESSIVE MYOCLONIC, 8; EPM8	Thin lips
51828	616230	EPILEPSY, PROGRESSIVE MYOCLONIC, 8; EPM8	Thin upper lip
51829	616230	EPILEPSY, PROGRESSIVE MYOCLONIC, 8; EPM8	Transferrin saturation may be low or normal
51830	616230	EPILEPSY, PROGRESSIVE MYOCLONIC, 8; EPM8	Triangular face
51831	616230	EPILEPSY, PROGRESSIVE MYOCLONIC, 8; EPM8	Tube feeding
51832	616230	EPILEPSY, PROGRESSIVE MYOCLONIC, 8; EPM8	Tube feeding
51833	616230	EPILEPSY, PROGRESSIVE MYOCLONIC, 8; EPM8	Tube-feeding
51834	616230	EPILEPSY, PROGRESSIVE MYOCLONIC, 8; EPM8	Uncoordinated gait
51835	616230	EPILEPSY, PROGRESSIVE MYOCLONIC, 8; EPM8	Unsteady gait
51836	616230	EPILEPSY, PROGRESSIVE MYOCLONIC, 8; EPM8	Variable manifestations
51837	616230	EPILEPSY, PROGRESSIVE MYOCLONIC, 8; EPM8	Variable severity

Processed genetic epilepsy data extracted from OMIM



Cellfie Transformation Rule



Ontology Prototype visualized using Protégé

Our ontology consists of the genetic epilepsy syndromes acting as the main classes and the clinical phenotypes and genotypes that cause them acting as the subclasses. Furthermore, in each subclass of clinical phenotype and genotype, all associated genetic epilepsy syndrome were also linked as shown in figure above by utilizing Protégé program. This represents the prototype/our first pass of our ontology building.

We subsequently attempted to create an more structured ontology where we would have 3 main classes representing the genetic epilepsy syndromes, clinical phenotypes and causal genotypes. Subclasses for each of these 3 main classes would be linked together through properties such as '*causedByPhenotype*', and '*causedByGenotype*'. We were unable to correctly create this desired ontology structure, despite repeated attempts using both Cellfie and even Python. However, we do believe that future work should look to model an ontology for genetic epilepsy syndromes as proposed above.

Problem Solving Method

From the processed genetic epilepsy syndrome data extracted from OMIM, we obtained a list of genetic epilepsy syndromes and their corresponding symptoms that are exhibited by patients.

After further processing to create a suitable dataset, we used this information to train a Bayesian Network to facilitate the diagnosis of genetic epilepsy syndromes. Specifically, the Bayesian model returns the most likely genetic epilepsy syndromes the patient is suffering from given the symptoms they are exhibiting.

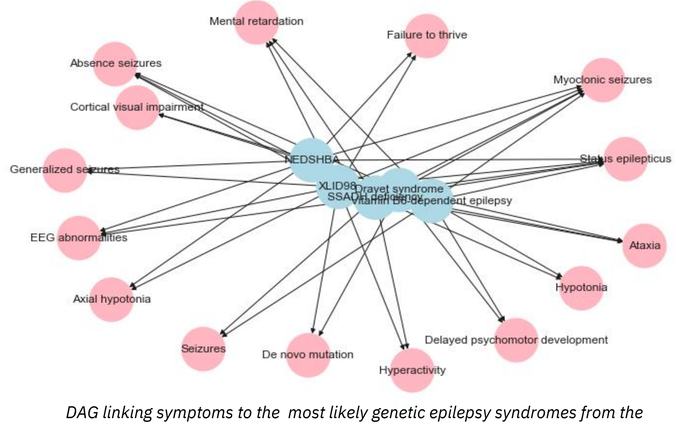
Case Study:

Q1: My patient has suffered from a *Myoclonic Seizure*, showing signs of *Status Epilepticus*. Which genetic epilepsy syndrome are they most likely to have?

Diagnosis	Probability
SUCCINIC SEMIALDEHYDE DEHYDROGENASE DEFICIENCY; SSADHD	0.122393
NEURODEVELOPMENTAL DISORDER WITH SEIZURES, HYPOTONIA; NEDSHBA	0.122393
INTELLECTUAL DEVELOPMENTAL DISORDER, X-LINKED 98; XLID98	0.122393
EPILEPSY, EARLY-ONSET, 4, VITAMIN B6-DEPENDENT; EPEO4	0.122393
DRAVET SYNDROME; DRVT	0.122393

Q2: Besides *Status Epilepticus* and *Myoclonic Seizures*, my patient additionally showed signs of *deterioration of cognitive function*. Based on this new evidence, which genetic epilepsy syndrome are they most likely to have?

Diagnosis	Probability
DRAVET SYNDROME; DRVT	***0.989332***
SUCCINIC SEMIALDEHYDE DEHYDROGENASE DEFICIENCY; SSADHD	0.00148772
NEURODEVELOPMENTAL DISORDER WITH SEIZURES, HYPOTONIA; NEDSHBA	0.00148772
INTELLECTUAL DEVELOPMENTAL DISORDER, X-LINKED 98; XLID98	0.00148772
EPILEPSY, EARLY-ONSET, 4, VITAMIN B6-DEPENDENT; EPEO4	0.00148772



DAG linking symptoms to the most likely genetic epilepsy syndromes from the case study above.

Domain Expert Evaluation

Domain Experts (board certified neurologists) at Stanford evaluated the quality of our Bayesian Model using the 10-point Likert scale, illustrated below.

1. To what extent do you believe the possible genetic epilepsy syndromes provided align with what you consider reasonable in the realm of a differential diagnosis? (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)

2. To what extent did the model present information that you had not previously considered or were unaware of? (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)

3. How accurate does the assigned probability of genetic epilepsy syndrome appear based on the Bayesian modeling? (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)

4. How helpful do you find the model's output in aiding the diagnostic workup of genetic epilepsy syndromes? (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)

Questions	Score from Domain Experts
To what extent do you believe the possible genetic epilepsy syndromes provided align with what you consider reasonable in the realm of a differential diagnosis?	8
To what extent did the model present information that you had not previously considered or were unaware of?	10
How accurate does the assigned probability of genetic epilepsy syndrome appear based on the Bayesian modeling?	7.5
How helpful do you find the model's output in aiding the diagnostic workup of genetic epilepsy syndromes?	8.5

As evident from the scores above, domain experts found the model predictions to be reasonable, accurate and helpful. They most appreciated the model's ability to present possible diagnoses they hadn't considered themselves, perhaps due to their rare occurrence in practice.

Takeaways

From our efforts in this project, we have compiled the following collective takeaways:

- ➔ Ontology and probabilistic reasoning based approaches like Bayesian Inference can truly improve diagnostic workups for underfunded diseases like epilepsy.
- ➔ The reliability and quality of the ontology and the problem solving method relies almost entirely on the quality of underlying data.
- ➔ Building ontologies at scale in an automated fashion is hard to achieve through Protege itself.
- ➔ Despite their simplicity, Bayesian network-based models can provide real benefit given their accuracy and ability to surface syndromes that could be left unconsidered.