

## Varient @ UofTHacks, Our Challenge

In Varient's mission to link people with rare diseases to treatment information for their specific condition, we discovered that the information gap among the rare disease community is, in part, caused by the lack of information standardization in the industry. With each hospital, clinic, and genetic testing agency documenting their data differently, one person's genetic report can look miles apart from the other, making it inefficient to extract meaningful information.

For our challenge to UofTHacks participants, we have provided a dataset for you to come up with an image recognition algorithm that can extract key information from a given genetic report. **For starters, try to extract the gene mutation reported.** Here are the details:

- We have named each file with the gene mutation of interest
- There are numbers after the file name with '\_' in front, they are not part of the gene name, they are only used to remove name duplicates
- Given the name of a gene of interest, identify the gene of interest on the report
- We accept any method of labelling the identified gene on the report
- If the labelled gene cannot be identified, please report that too.

Here's our dataset:

<https://drive.google.com/drive/folders/1G4BkT2SLpRSC2rjwLcvmtYA4Q0osriCj?usp=sharing>

Open a mentor ticket or find us under the Varient channel in discord if you have any questions!