**S1 Text. Machine learning algorithm for mining disease-mutation relationships:**

We briefly discuss the machine learning and text mining approach we developed to extract disease-mutation relationships from PubMed abstracts. As shown in Supplementary Figure 1 (below), using the tmVar and DNorm entity recognition tools for mutation and disease mentions respectively, we determine an association (yes or no) by computing six quantitative features from the text in the abstract. These 6 features are constructed to capture various information types such as statistical information about frequency, inter-entity distances, and text sentiment and its strength. We use a pre-trained (on a human annotated benchmark dataset) machine learning model to make decisions about associations using the six feature set.



Supplementary Figure 1: Flowchart of our disease-mutation association mining algorithm.