With the advent of next-generation sequencing and the completion of the human genome there has been an explosion of sequencing data. One problem with these new technologies, however, is that while they are extremely sensitive they are also error prone and subject to biases inherent in particular DNA sequences. In addition, the lack of standardized protocols to isolate DNA leads to biases from the choice of technique. As a result, separating true signals from background signals can be challenging and requires a significant investment of time in validating sequencing data using orthogonal methods. The National Center for Biotechnology Information (NCBI) warehouses this next-generation sequencing data for unrestricted use by other researchers. I will to explore the underlying structure of some of this data with unsupervised classification techniques. In addition, I will perform supervised classification on a subset of previously validated datasets in combination with features from the genomic context (e.g. distance to annotated genes) and the sequence itself (e.g. local repetitiveness or distance to chromosome ends) to classify unvalidated sequencing results as true positives or false positives. I hope that this work will reveal patterns that can help prioritize results from these experiments and increase efficiency when it comes to validation of findings.