

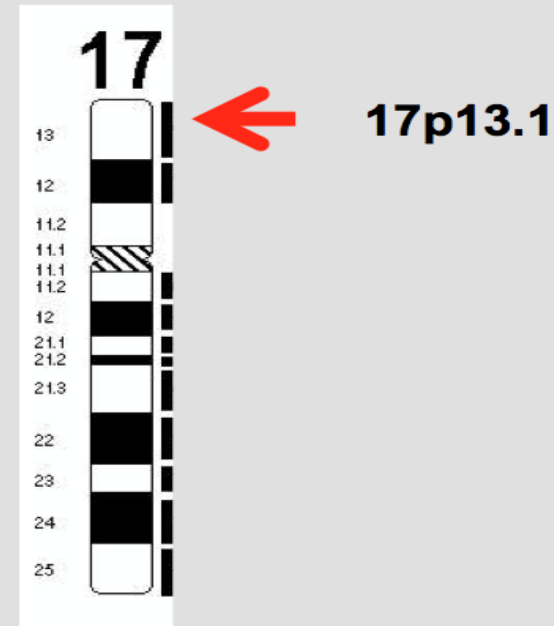
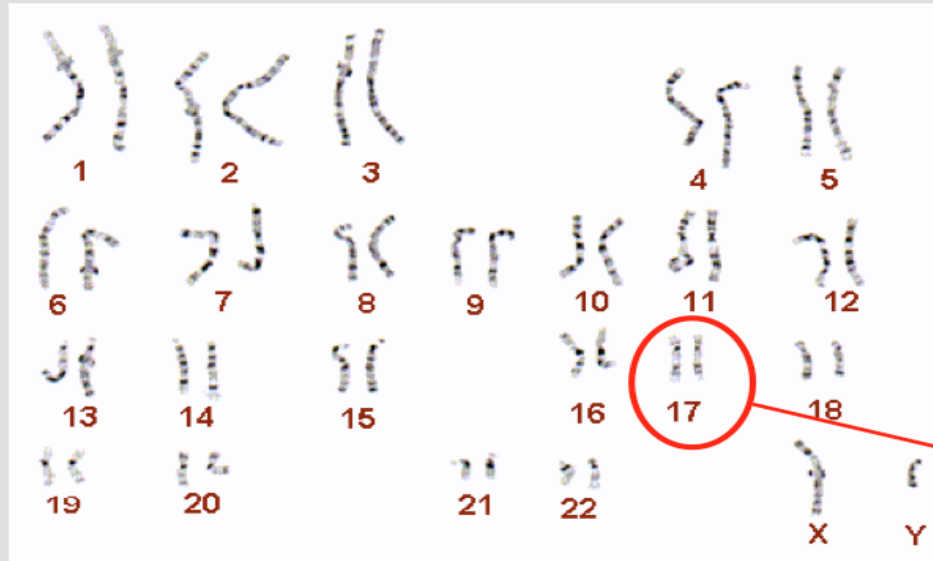


Exploring the Oncogenic Impact of The **TP53** Gene Variants and Classifying the Pathogenicity¹

By Alison Vaz

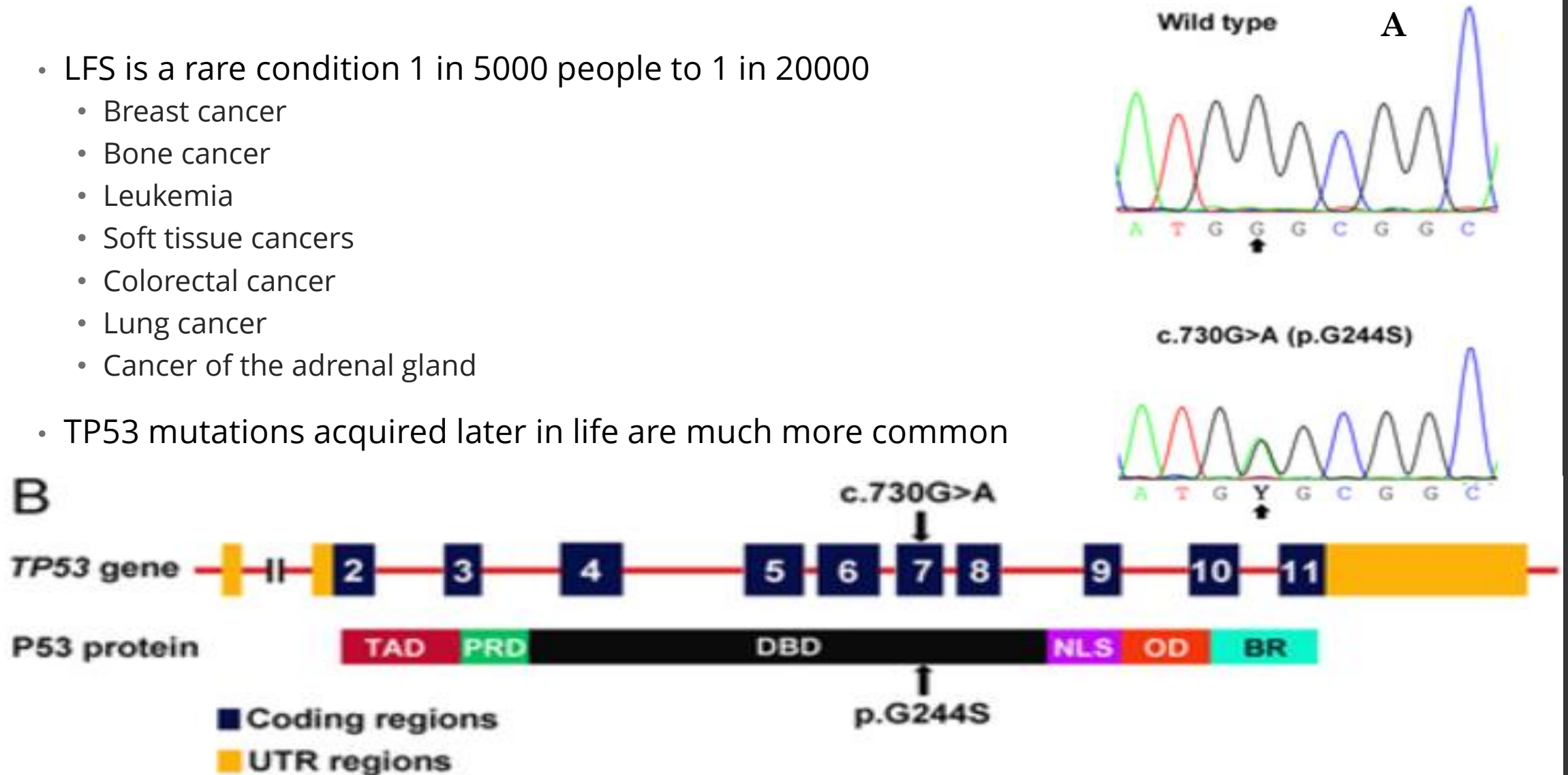
What Is the TP53 Gene?

LOCALIZATION OF THE TP53 GENE



If You Carry a TP53 Gene Mutation, What Cancers Are You at Risk For?

- LFS is a rare condition 1 in 5000 people to 1 in 20000
 - Breast cancer
 - Bone cancer
 - Leukemia
 - Soft tissue cancers
 - Colorectal cancer
 - Lung cancer
 - Cancer of the adrenal gland
- TP53 mutations acquired later in life are much more common

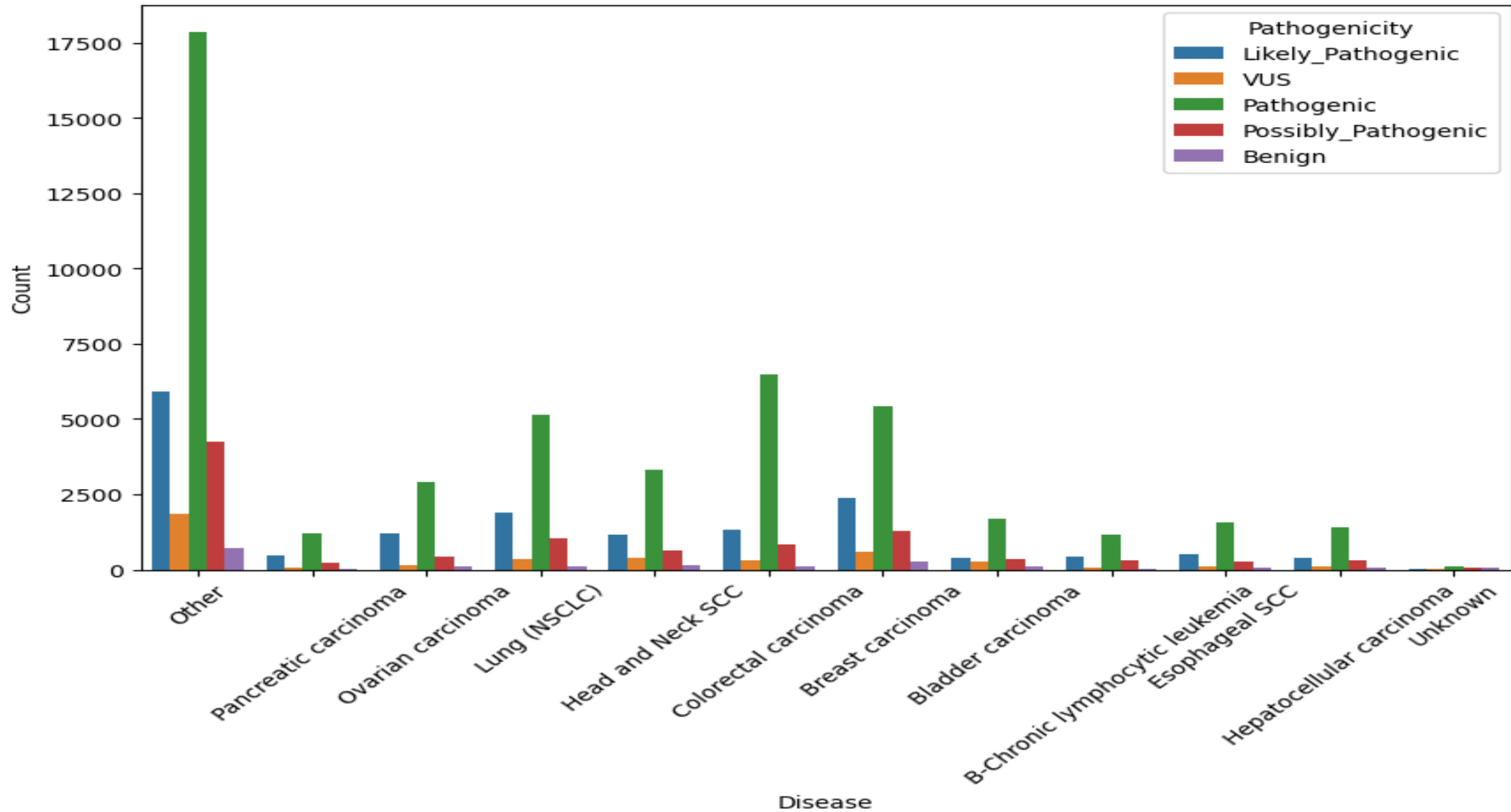


What is in my dataset?

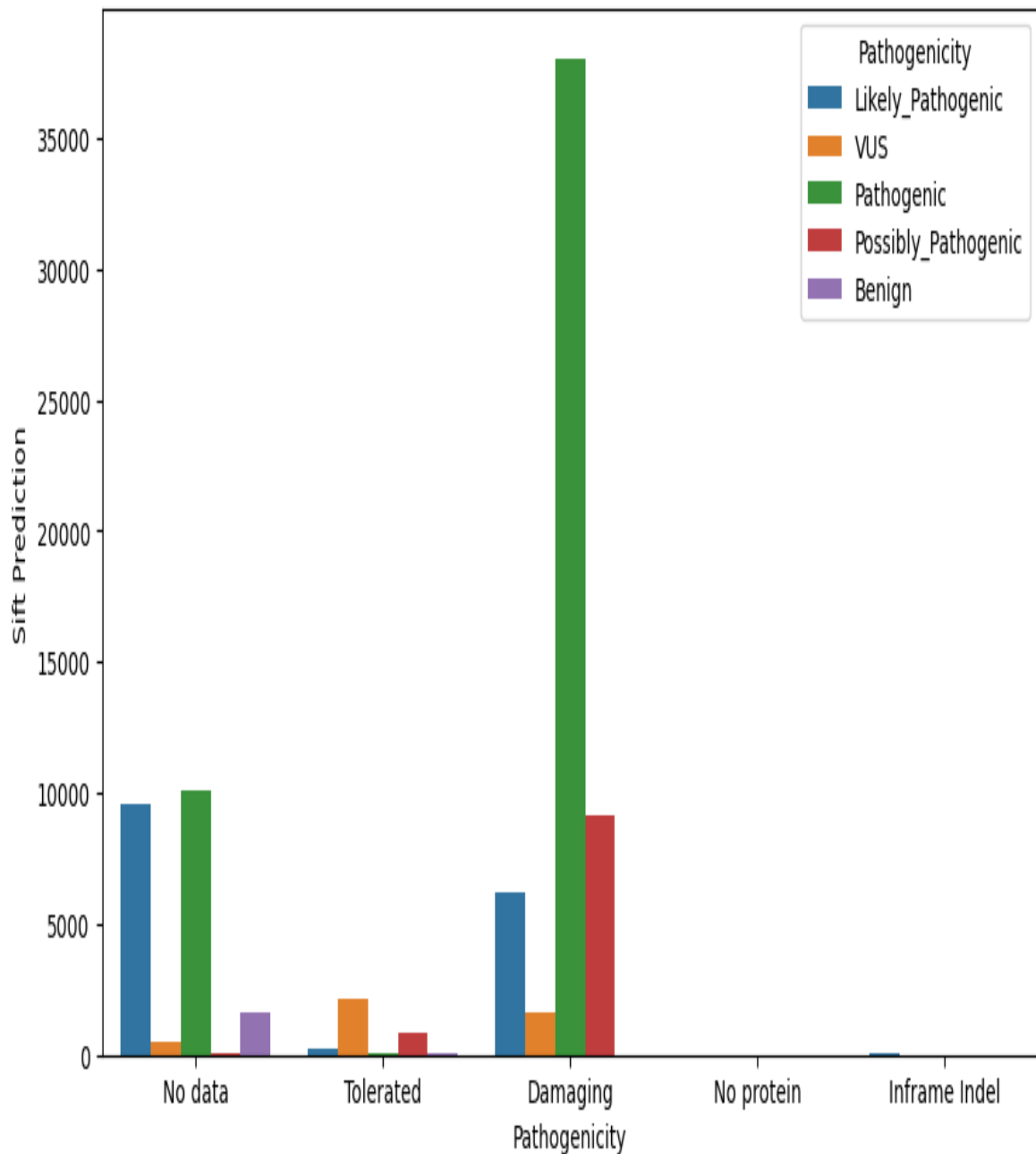
Pathogenicity (Target)	TP53 variants: 'pathogenic', 'likely pathogenic', 'uncertain significance' (VUS), 'likely benign'
Exon:intron_start	Start position of exon or intron
Exon:intron_End	End position of exon or intron
Genome_base_coding	Unmutated nucleotide in the genome
Mutant_Allele	Mutated nucleotide
Mutation_Type	Insertion, Deletion, Indel, SNV
Ins_Size	Size of insertion in mutation
Del_Size	Size of deletion in mutation
Disease	Associated cancer
Tumour_ Repetition	Denotes the total number of mutations associated with this mutant in a single tumour
Sift_Prediction	Predicted function effect of the SIFT algorithm
Provean_Prediciton	Predicted function effect using the PROVEAN algorithm

And much more...

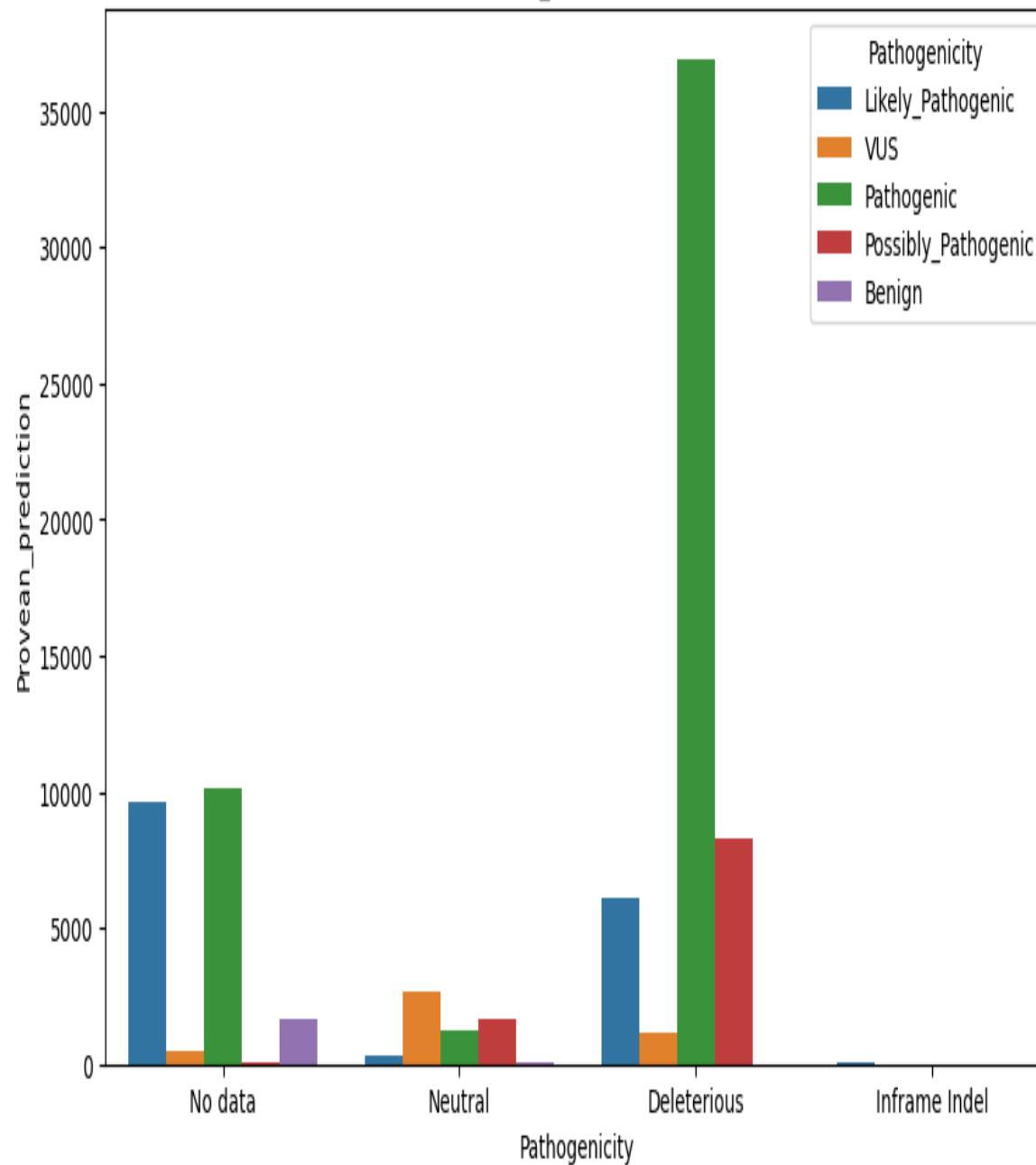
Distribution of Disease by Pathogenicity



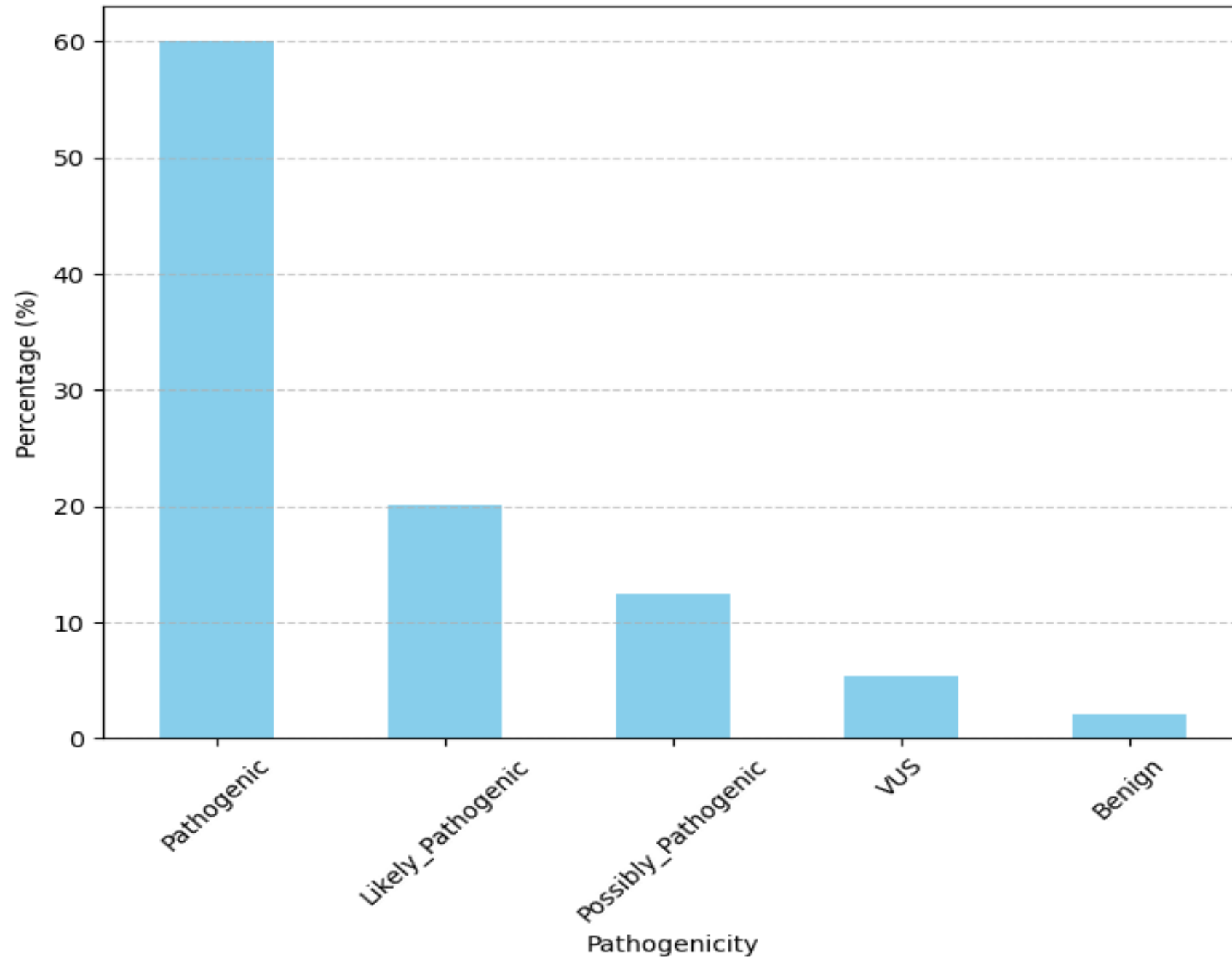
Countplot of Sift Prediction by Pathogenicity



Countplot of Provean_prediction by Pathogenicity



Class Imbalance of Pathogenicity Column



Baseline Modeling

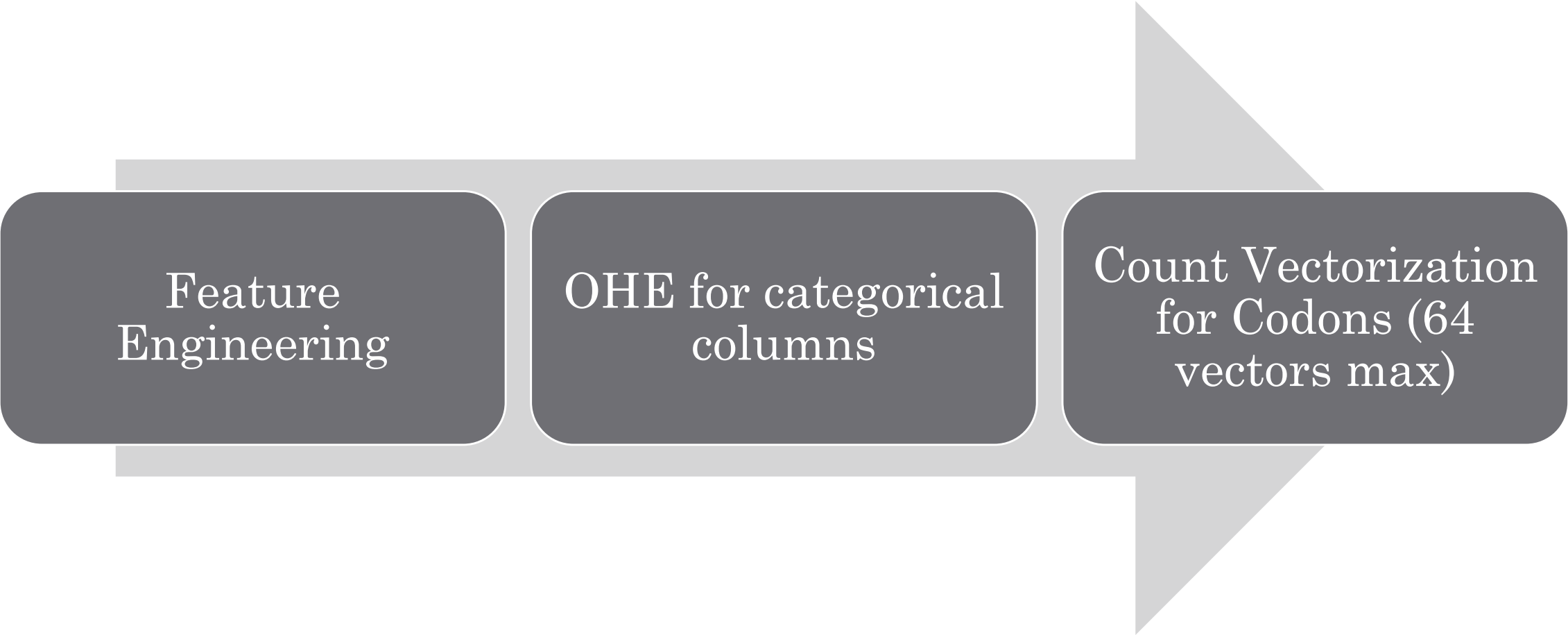
Logistic
Regression

Test Accuracy =
72.08%

Random
Forest

Test Accuracy =
99.39%

Next Steps



Feature
Engineering

OHE for categorical
columns

Count Vectorization
for Codons (64
vectors max)

REFERENCES

- The TP53 Website
- <https://www.everydayhealth.com/cancer/risk-genes-what-you-need-know-about-tp53/>
- https://www.researchgate.net/figure/G244S-mutation-of-the-TP53-gene-A-Sanger-sequencing-of-codons-727-735-of-wild-type_fig2_292212402