The choice of coding design was driven by several key considerations. When I initially encountered the data set, my primary concern was how to handle the extensive 22,277 columns of data. Additionally, upon closer examination of the class methods, I determined that using a dictionary structure would offer the most practical solution, enabling easy access to specific values associated with a given key.

To implement this design, I created a "LiverData" class in which attributes were designated for sample names, cancer states, and normal state gene values. Subsequently, I instantiated instances of this class as the values within the dictionary, with each gene name serving as a key. This structure resulted in a dictionary containing each gene name and its corresponding 377 class objects.

For the purpose of calculating differential equations, I compiled lists of gene values associated with both the "hcc" and normal states. The formula used for these calculations was: log2(mean(hcc_state_gene_values) / mean(normal_state_gene_values)). These differential equations were computed for each gene and subsequently stored within the dictionary as values, with the gene name as the corresponding key. This approach ensured that one could easily access the differential equations for any specific gene by using its name.

It's important to note that I made the assumption that the data was distributed normally. That’s why I utilized statistical functions from SciPy to calculate p-values. I did the t-test to see the significance difference between hcc state gene values and normal state genes values.

To further enhance the visual representation of the results, I took the negative natural logarithm of these p-values and plotted a scatter plot. In this graph, the x-axis represented the differential expression values, while the y-axis displayed the associated p-values, creating a comprehensive visualization of the data.