Dataset Size vs Minimum Disease Sample Threshold

Minimum samples per disease	Number of diseases	Total samples retained
100	443	44300
200	361	72200
300	304	91200