**Exploratory Data Analysis**

Chart, scatter chart

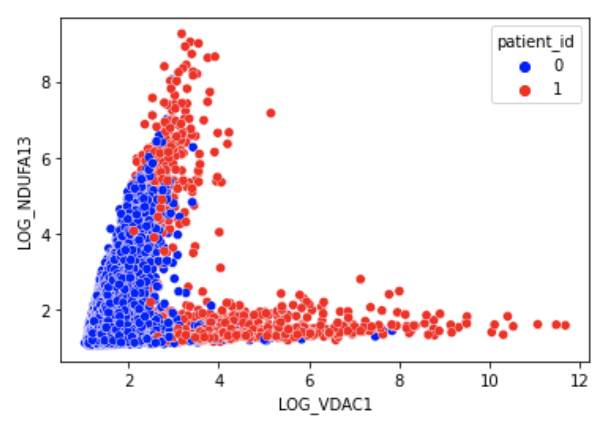
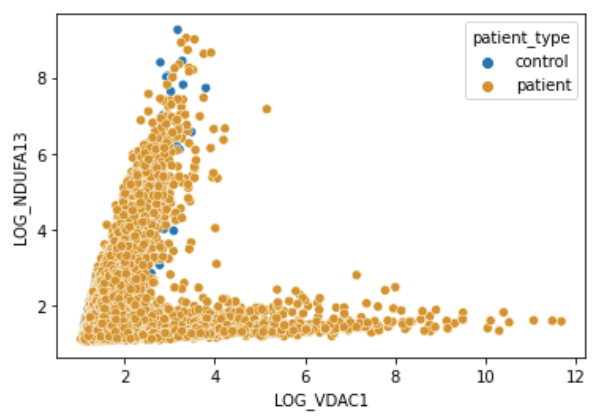
Description automatically generated

Chart, scatter chart

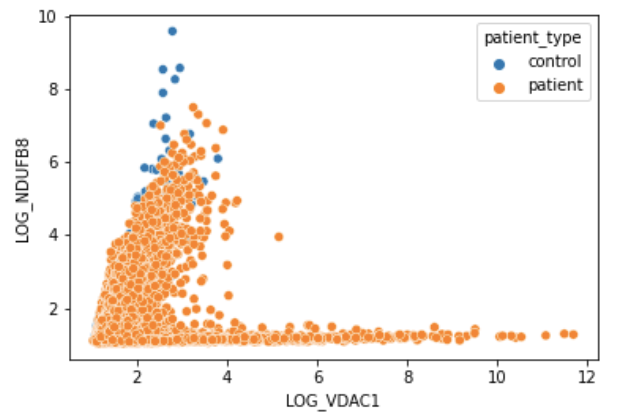
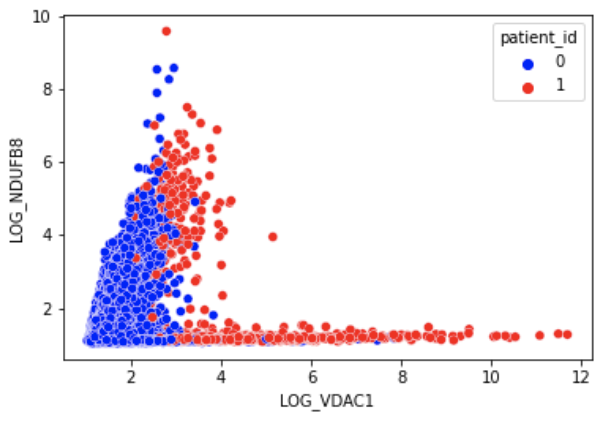
Description automatically generated

The two graphs above show no clear separation between the patient and the control fibres for those proteins.

Raw data K-means



Raw data K-means



The K-means appears to be producing a similar result to the raw data.

K-Mean results

|  |  |  |
| --- | --- | --- |
| **Individual** | **Healthy Fibre** | **Disease Fibre** |
| C01 | 148 | 0 |
| C02 | 289 | 0 |
| C03 | 102 | 29 |
| P01 | 69 | 268 |
| P02 | 24 | 208 |
| P03 | 1361 | 0 |
| P04 | 733 | 146 |
| P05 | 1866 | 12 |
| P06 | 538 | 270 |
| P07 | 581 | 174 |
| P08 | 531 | 97 |
| P09 | 892 | 54 |
| P10 | 225 | 377 |

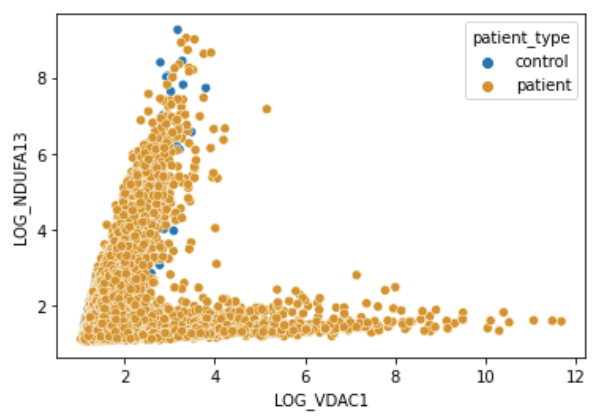
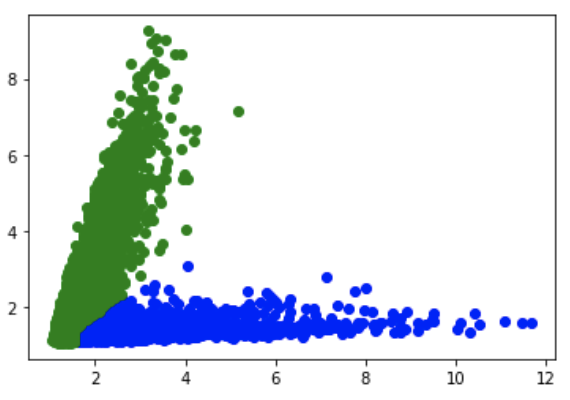
The algorithm predicted the control to all have mainly healthy fibres, whereas, for patients there is a mixture of healthy and disease fibres. This led us to investigate whether the type of disease had an impact on the number of the fibres.

|  |  |  |
| --- | --- | --- |
| **Disease Type** | **Healthy Fibre** | **Disease Fibre** |
| **Control** | 539 | 29 |
| **CI** | 93 | 476 |
| **Deletion** | 2094 | 146 |
| **MT-TL1** | 2985 | 456 |
| **MT-TG** | 531 | 97 |
| **MT-TE** | 892 | 54 |
| **MT-TW** | 225 | 377 |

The CI mutation disease presents with much higher numbers of disease fibres in comparison to the healthy fibres. The Deletion, MT-TL1 and MT-TE diseases all have a higher healthy fibre count than disease. MT-TW seem to have a similar number of healthy and disease fibres.

**GMM**

Raw data GMM



Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generatedRaw data GMM

The GMM graphs show a similar shape, for both proteins, in comparison to the raw data graphs. The GMM for NDUFB8 does not appear to split the data into two clusters like it does for the NDUFA13 graph.