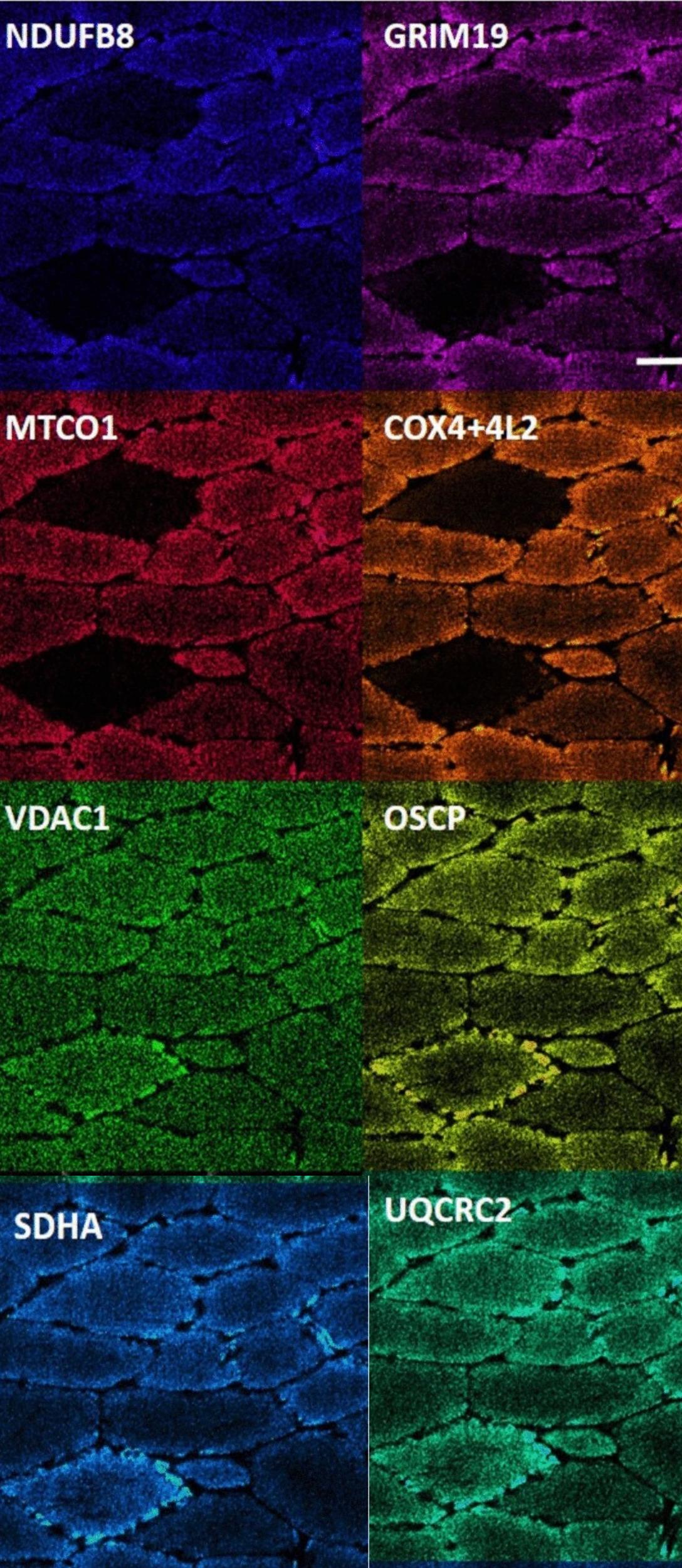


Introduction



The Welcome Centre for Mitochondrial Research obtained data from 13 anonymous individuals of which 10 had varying types of mitochondrial disease and 3 did not, whom we used as control. Samples were obtained and used to create the dataset used in this project via the use of imaging mass cytometry (IMC) to obtain information about the protein expressions of the patient fibres. There were 6 different disease across all 10 patients and 9000 fibres in the whole dataset. The proteins NDUFB8 and NDUFA13 were run in the k-means and gaussian mixture model (GMM). This is because they form part of complex I (CI) of the respiratory chain (RC), used for energy production. Complex I is known to be deficient in those with mitochondrial disease and therefore would expect those two proteins to be absent in the CI patient fibres, and present in control fibres.

The aim of the study

This project aims to design a model to accurately classify fibres into one of two groups: healthy or RC deficient. Followed by identifying how these classifications relate to the different mitochondrial diseases. This will involve carrying out two clustering algorithms to identify whether either of one of them is able to produce two clusters: one for each classification.

FIBRE CLASSIFICATION METHOD

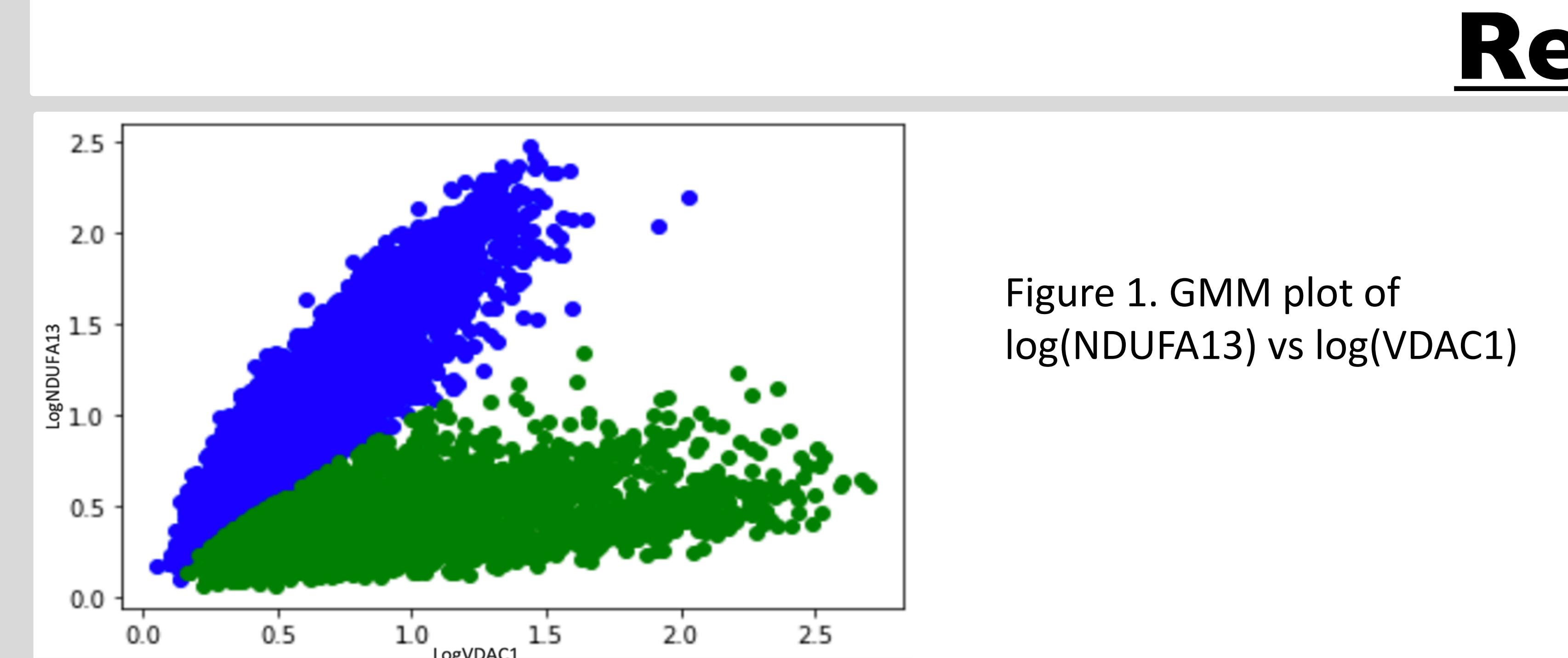
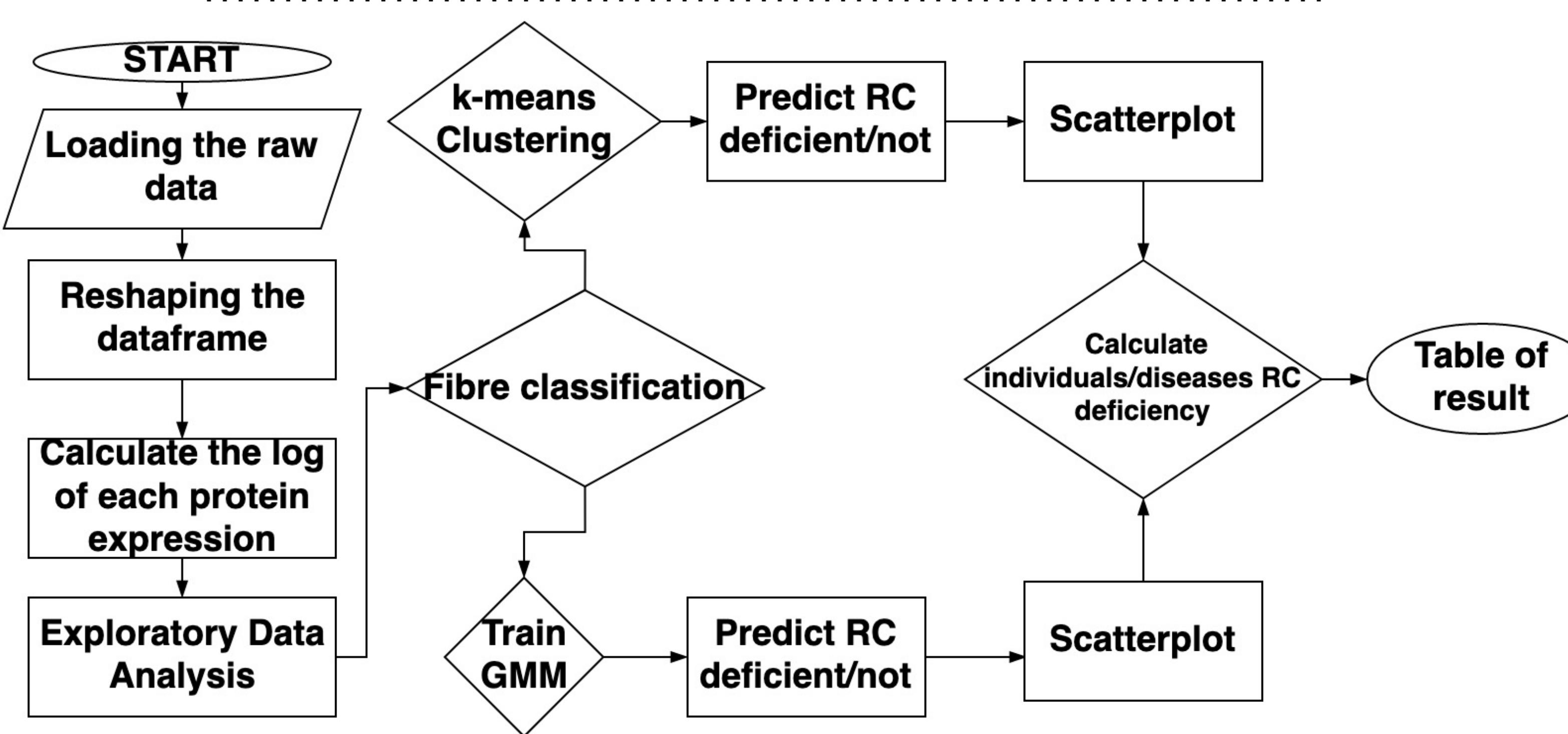


Figure 1. GMM plot of log(NDUFA13) vs log(VDAC1)

The GMM scatterplot produced two clusters with a clear V-divide for the two proteins as expected, blue for the control and green for the patients. This was not observed in the k-means plot highlighting that the GMM was a more suited model to this dataset.

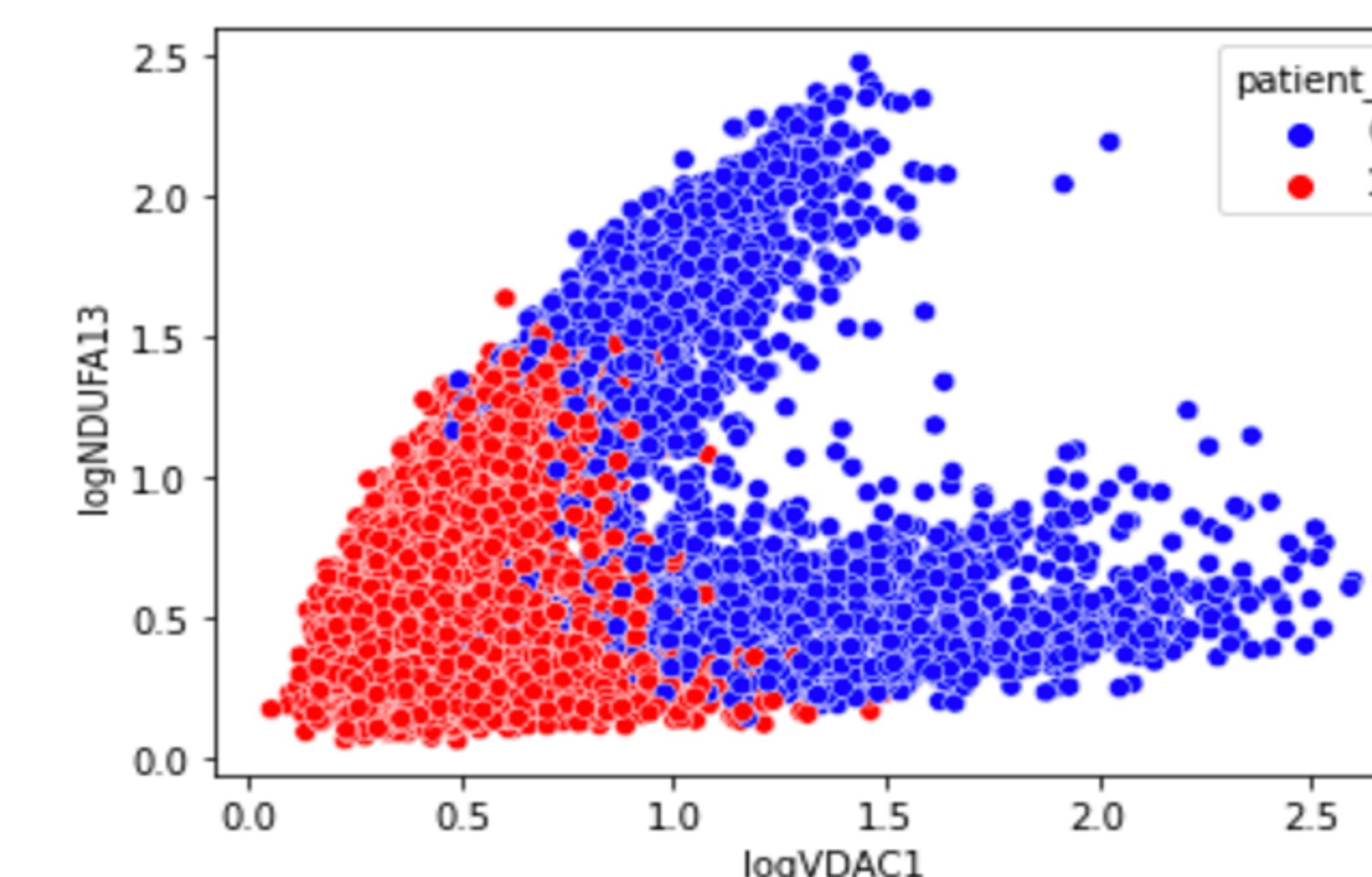


Figure 2. k-means plot of log(NDUFA13) vs log(VDAC1)

Individual	Proportion of Disease Type	Proportion of RC deficient Fibres (%)
C01	1.4%	
C02	0%	Control
C03	80.9%	
P01	97.6%	CI
P02	98.7%	
P03	0.8%	Deletion
P04	84.8%	
P05	20.1%	
P06	76.4%	MT-TL1
P07	49.5%	
P08	54.8%	MT-TG
P09	50.2%	MT-TE
P10	80.9%	MT-TW

Table 1. Proportion of RC Deficient Fibres in each patient and in each disease from the k-means result.

Results

The GMM predicted the control as having approximately 0% of RC deficient fibres.

The fibres of the patients with the deletion variant were predicted to have a low RC deficient and this was seen across both patients as expected. The CI variant was found to have 100% of the fibres as RC deficient as expected.

Individual	Proportion of RC deficient Fibres (%)	Disease Type	Proportion of RC deficient Fibres (%)
C01	0%		
C02	0.3%	Control	0.2%
C03	0%		
P01	100%	CI	100%
P02	100%		
P03	2.6%	Deletion	1.5%
P04	3.4%		
P05	21.4%		
P06	9.5 %	MT-TL1	28.2%
P07	65.3%		
P08	83.6%	MT-TG	83.6%
P09	37.7%	MT-TE	37.7%
P10	80.9%	MT-TW	80.9%

Table 2. Proportion of RC Deficient Fibres in each patient and in each disease from the GMM Result.

Conclusion

Comparison of the GMM and k-means models highlight that GMM was better suited to this dataset. This was reinforced by the scatterplots which showed GMM to produce a better split of the clusters.

The expectation of the CI fibres to be 100% RC deficient was found in both clustering models highlighting that this assumption was correct.

MT-TL1 was found to have varying proportions of RC deficient fibres in all the patients highlighting the likelihood of an alternative cause for that version of the mitochondrial disease.

Future Work

Looking at MT-TL1, each patient in that same category had varying proportions of RC deficient fibres for the two algorithms. It would be interesting to take this study further to identify what other attributes could influence the classification of an individual into this category considering their varying protein contents. This might include looking into deficiency of other proteins produced by mitochondrial tRNA.

Thanks & appreciation

Deborah Ngongo, Stephen McGough, Matt Forshaw, Amy Vincent, Connor Lawless & Atif Khan

Contact

E: frestie_ngongo@hotmail.com LinkedIn: Frestie Ngongo GitHub: FNgongo