APPENDIX B – DISCRIMINATIVE EVENT BASED MODELLING

DEBM: Gaussian mixture modelling

DEBM uses Gaussian mixture modelling to transform biomarker values to posterior probabilities of them being abnormal. This is done by assuming the probability density functions of normal and abnormal values are represented by Gaussians $N(\mu_\simE, \sigma_\simE)$ and $N(\mu_E, \sigma_E)$ respectively, where the occurrence of the biomarker abnormality event is denoted by $E$ and the absence of such an event is denoted by $\simE$.

Gaussian mixture modelling is an optimisation task to estimate these normal and abnormal Gaussians as well as the mixing parameter based on maximum log-likelihood, where the log-likelihood for biomarker $B$ is computed as the summation over all $GRN$ mutation carriers in the dataset as follows:

$$L_B = \sum_{j \in \text{Carriers}} \log f(B_j)$$

Figure B.1: Illustrations of the Gaussian probability density functions for normal and abnormal values of biomarker $B$

Here, the likelihood $f(B)$ is computed as follows:

$$f(B) = \theta_\simE p(B|\mu_\simE, \sigma_\simE) + \theta_E p(B|\mu_E, \sigma_E).$$
Where $\theta_{-E} + \theta_E = 1$, and the mixing parameters $\theta_{-E}$ and $\theta_E$ show the relative proportions of the two Gaussians in the dataset. The abnormal Gaussian is initialized using the mean and standard deviation of the symptomatic subjects, while the normal Gaussian is initialized using the non-carriers. Since non-carriers are healthy controls, we fix $\mu_{-E}$ and $\sigma_{-E}$ to their initialized values and only optimize the remaining parameters in the Gaussian mixture model. The mixing parameter and the Gaussian parameters are optimized alternately until convergence as detailed previously,[1].

For imaging-biomarkers with left and right counter parts, we propose a novel modification to the Gaussian mixture model optimization called Siamese Gaussian mixture model (Siamese GMM). We propose to jointly optimize the parameters of these biomarkers, by taking advantage of symmetry in the brain. The log-likelihood for the joint optimization for the imaging biomarkers $I^L$ and $I^R$ is given below:

$$L_I = \sum_{j \in \text{Carriers}} \log f(I^L_j) + \log f(I^R_j)$$

where $f(I^L_j)$ and $f(I^R_j)$ are expressed mathematically as:

$$f(I^L_j) = \theta_{-E}^L \rho(I^L_j | \mu_{-E}, \sigma_{-E}) + \theta_E^L \rho(I^L_j | \mu_E, \sigma_E)$$

$$f(I^R_j) = \theta_{-E}^R \rho(I^R_j | \mu_{-E}, \sigma_{-E}) + \theta_E^R \rho(I^R_j | \mu_E, \sigma_E)$$

$\theta_{-E}^L + \theta_E^L = 1$ and $\theta_{-E}^R + \theta_E^R = 1$. The mixing parameters $(\theta_{-E}^L, \theta_E^L, \theta_{-E}^R, \theta_E^R)$ and the abnormal Gaussian parameters $(\mu_E, \sigma_E)$ are again optimized alternately until convergence,[1]. This joint optimization of the left and right counter parts by sharing the normal and abnormal Gaussians reduces the number of parameters to be optimized, and thus improves the robustness. In case of asymmetrical atrophy patterns, where one of the biomarkers is stronger than the other, the joint optimization also helps in making the GMM more stable for the weaker biomarker.
Figure B.2. Correlation of disease severity (as estimated by non-imaging DEBM using cross-validation) with years since onset and FTD-CDR-SOB. The 2D scatter plots in figures A and C show the correlations of disease severity with years since onset, for symptomatic nfvPPA and bvFTD subjects respectively. The 2D scatter plot in figures B and D show the correlations of disease severity with FTD-CDR-SOB. The plot on top of each subfigure shows the probability density function of the disease stages. The plots on the right of figures A and C show the probability density functions of years since symptom onset. The plots on the right of figures B and D show the probability density function of FTD-CDR-SOB.
Figure B.3. Cascade of non-imaging biomarker changes in nfvPPA (A) and bvFTD (B) subjects along with the uncertainty associated with it. The biomarkers are ordered based on the position in the estimated cascade. The color-map is based on the number of times a biomarker is at a position in 100 repetitions of bootstrapping.

References