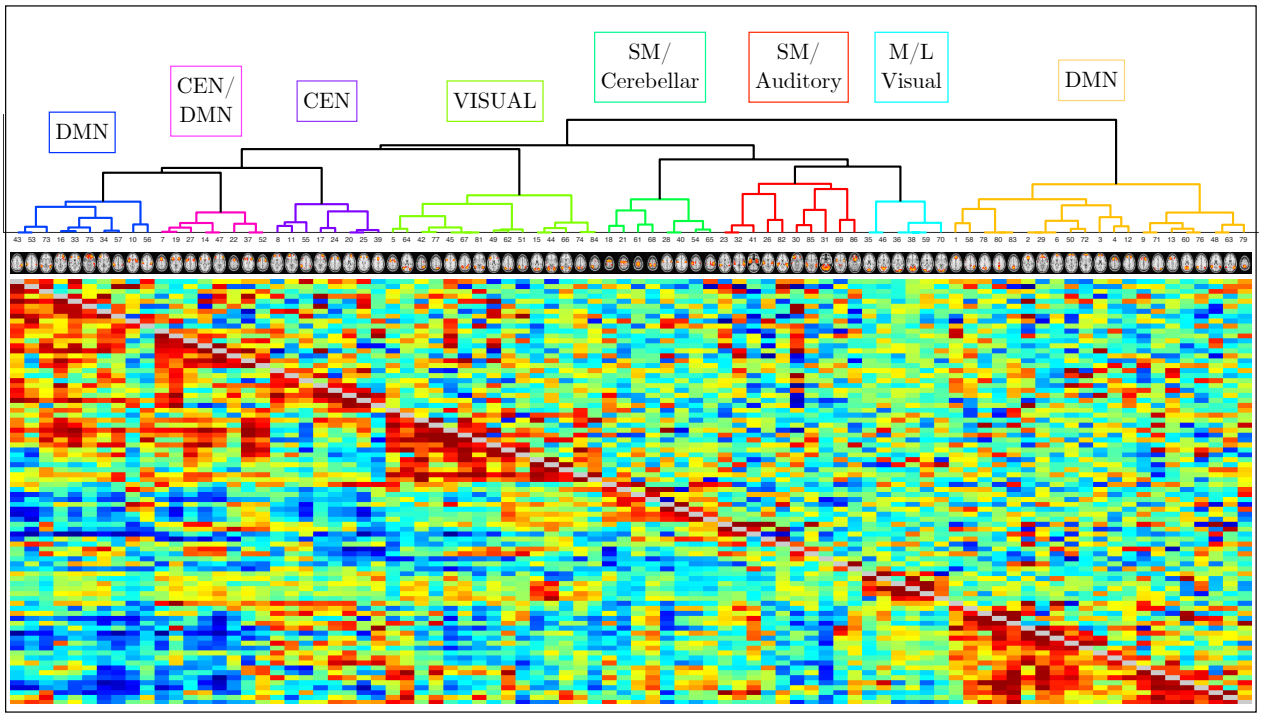
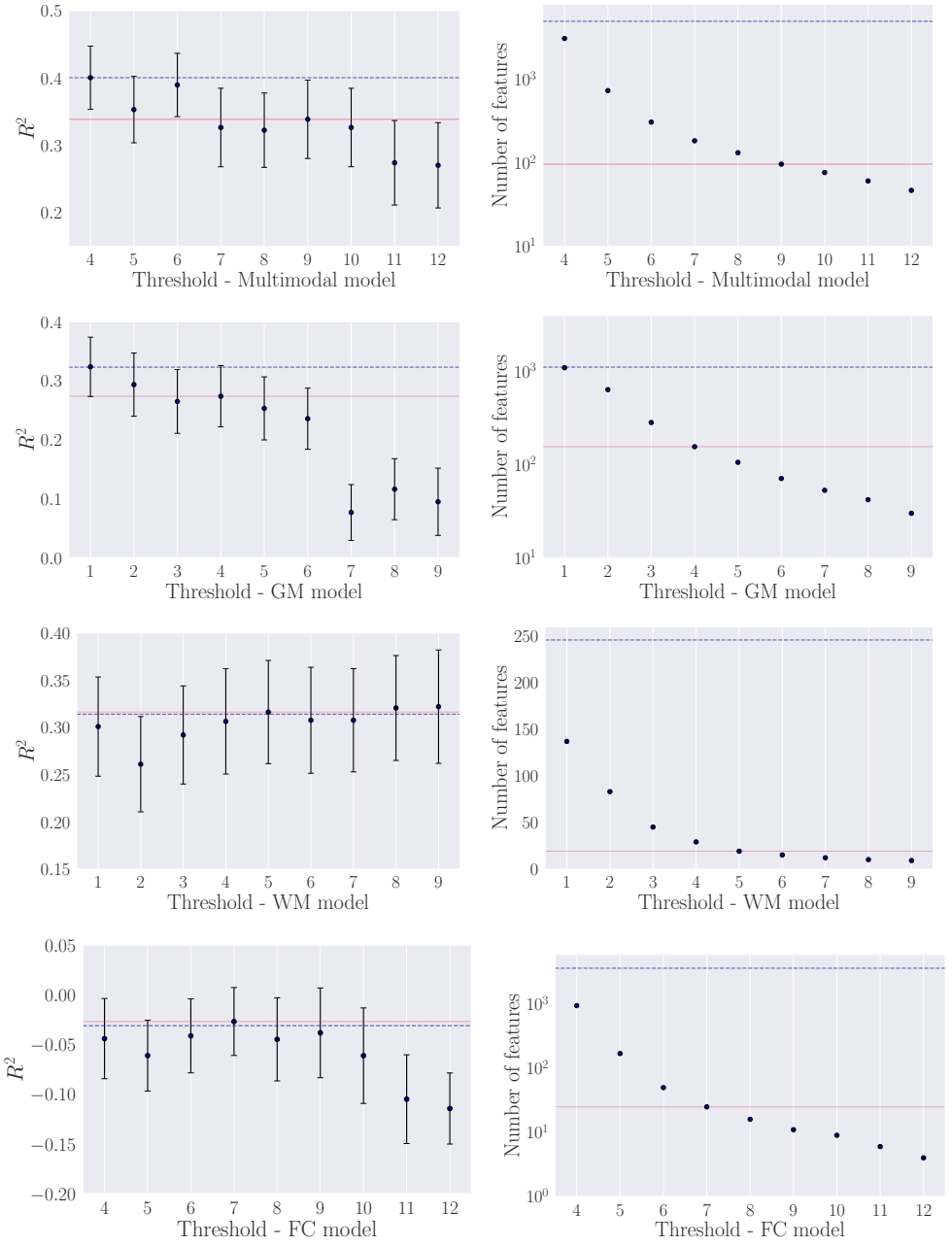
Supplementary Information

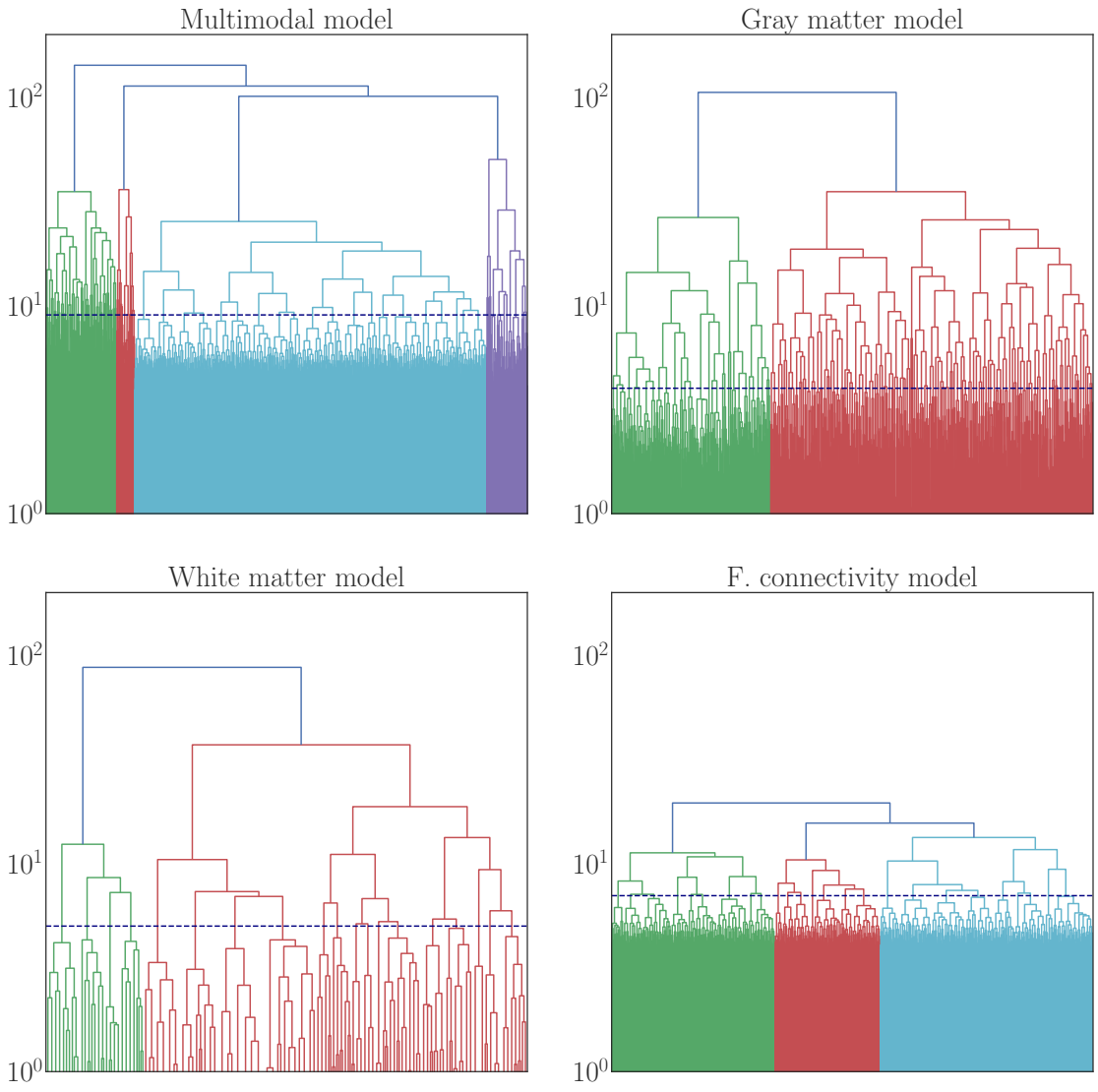


SI Figure 1: The resting-state functional connectome estimated for 671 participants of the Whitehall II MRI sub-study. High-dimensionality group ICA and network modelling were performed using the FSL-MELODIC and FSLNETS tools. Z statistics for the full correlation (below the diagonal) and partial correlation (above the diagonal) were computed for the 86 nodes visualized at the top of each column. The nodes were reordered according to a hierarchical clustering of the full correlation matrix. The eight clusters overlapping with commonly observed resting-state networks are and labelled at the top of the figure. DMN = default mode network; CEN = central executive network; SM = Sensorimotor; M = medial; L = lateral.

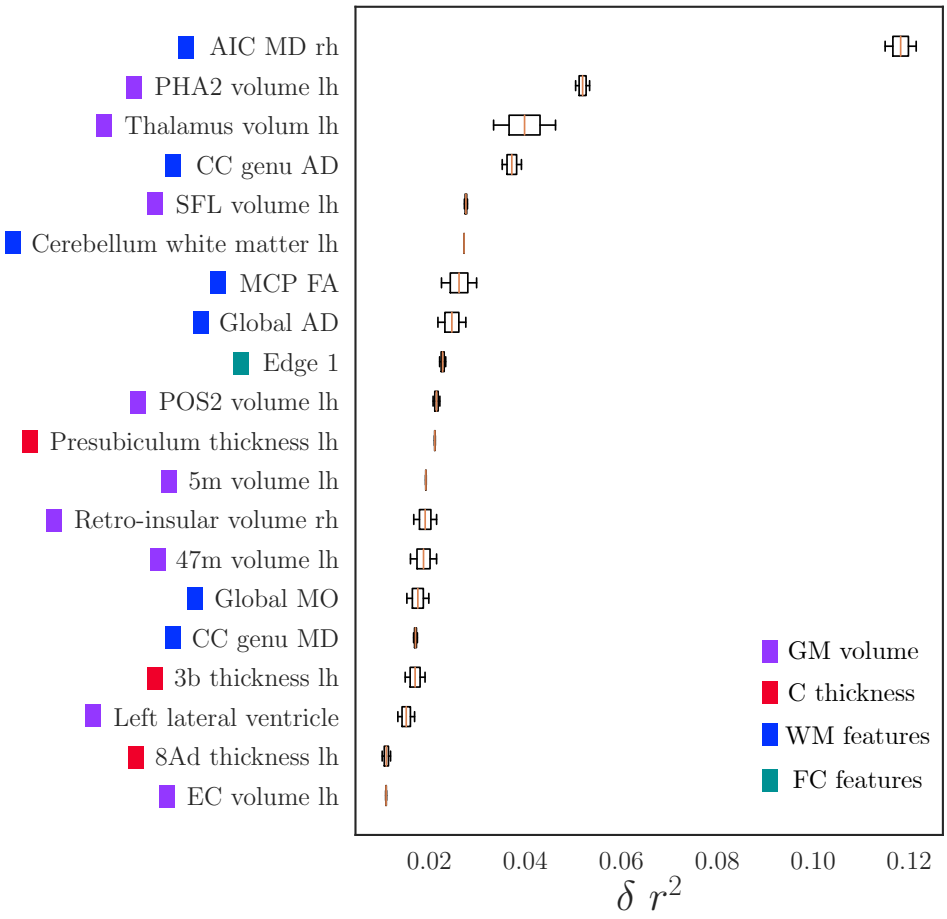
SI Figure 2: The brain regions for each of the eight clusters based on the resting-state functional connectome shown in SI Figure 1. DMN = default mode network; CEN = central executive network; SM = Sensorimotor; M = medial; L = lateral.



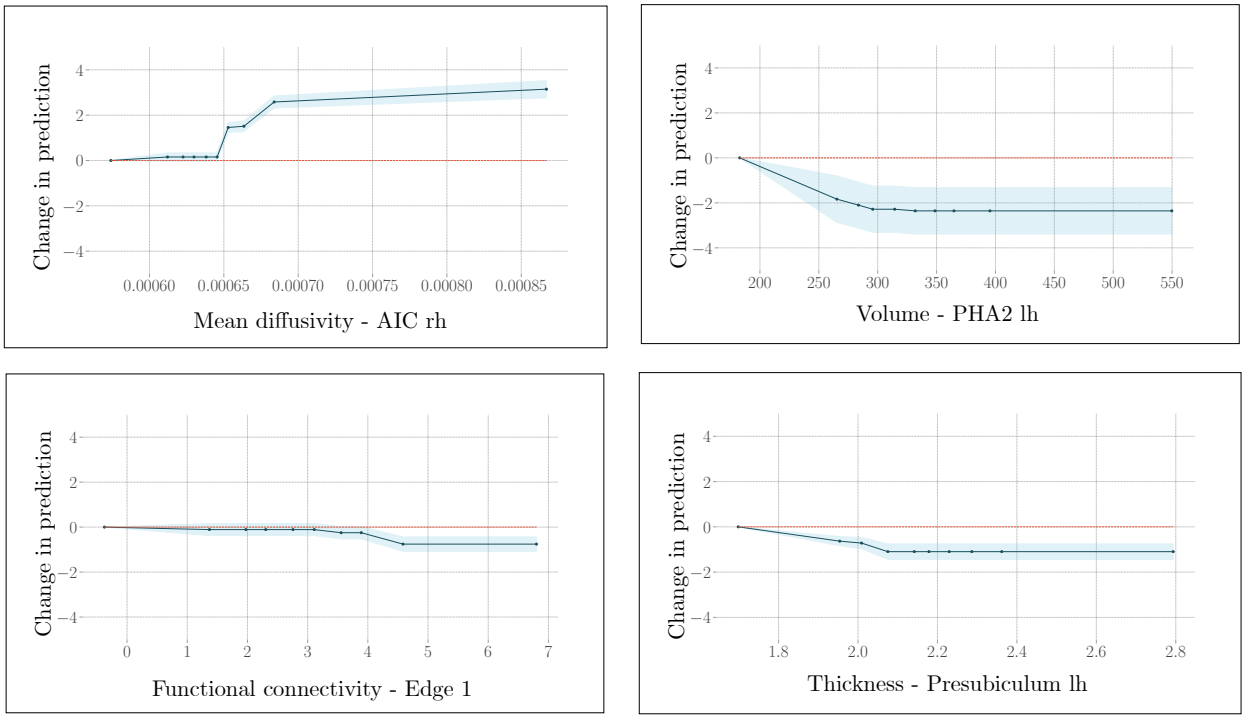
SI Figure 3: Left plots: R^2 values for each cluster threshold. The error bars indicate the standard error on the R^2 values. Blue, dotted lines indicate the R^2 for the full models with no threshold. Red lines indicate the R^2 for selected thresholds. Right plots: The y axis shows the number of features included with each threshold, with logarithmic scales for the multimodal, gray matter (GM), and functional connectivity (FC) models, and linear scale for the white matter (WM) model. Blue, dotted lines indicate the number of features for the full models with no threshold. Red lines indicate the number of features for each selected threshold. The left-most threshold value is based on the first value excluding some features. For example, if the threshold is set to ¡ 4 for the FC model, all features are included.



SI Figure 4: Dendrograms showing the relationships between features based on hierarchical clustering on the Spearman rank-order correlations. For each model, the colours represent distinct clusters of features, which are grouped together according to their degree of co-linearity. Clusters appearing at higher y-values (logarithmic scale) contain variables that are related but less co-linear. The horizontal blue dotted lines indicating the selected threshold for each model; 9 for the multimodal model, 4 for the gray matter model, 5 for the white matter model, and 7 for the functional connectivity model.



SI Figure 5: The top 20 features in the multimodal model run with 50/50 train-test split, ranked by permutation feature importance. The $\delta\ r^2$ values represent the decrease in model performance when the single feature value is randomly shuffled. Higher $\delta\ r^2$ values indicate that the feature contributes more to the model accuracy. Note that this feature ranking should not be interpreted as a list of imaging-derived features that are most related to age, as the clustering procedure excludes correlated features that could be equally age-dependent. AIC = Anterior Limb of Internal Capsule; AD = Axial Diffusivity; C = cortical; CC = Corpus Callosum; EC = Entorhinal Cortex; Edge 1 = functional connectivity between Anterior Cingulate Gyrus and Superior Frontal Gyrus; FA = Fractional Anisotropy; FC = functional connectivity; GM = gray matter; lh = left hemisphere; MCP = Middle Cerebellar Peduncle; MD = Mean Diffusivity; MO = Mode of Anisotropy; rh = right hemisphere; WM = white matter. PHA2, SFL, POS2, 5m, 47m, 3b, and 8Ad are defined in SI_Glasser_2016.pdf. The cortical regions of interest are based on the Glasser parcellation [35]; the subcortical regions of interest are based on the FreeSurfer atlas [34]; the white matter tracts are based on the CBM-DTI-81 White-Matter Labels Atlas [40, 41]; functional connectivity networks are based on an Independent Component Analysis [44] as described in Section 2.2.3 in the main manuscript.



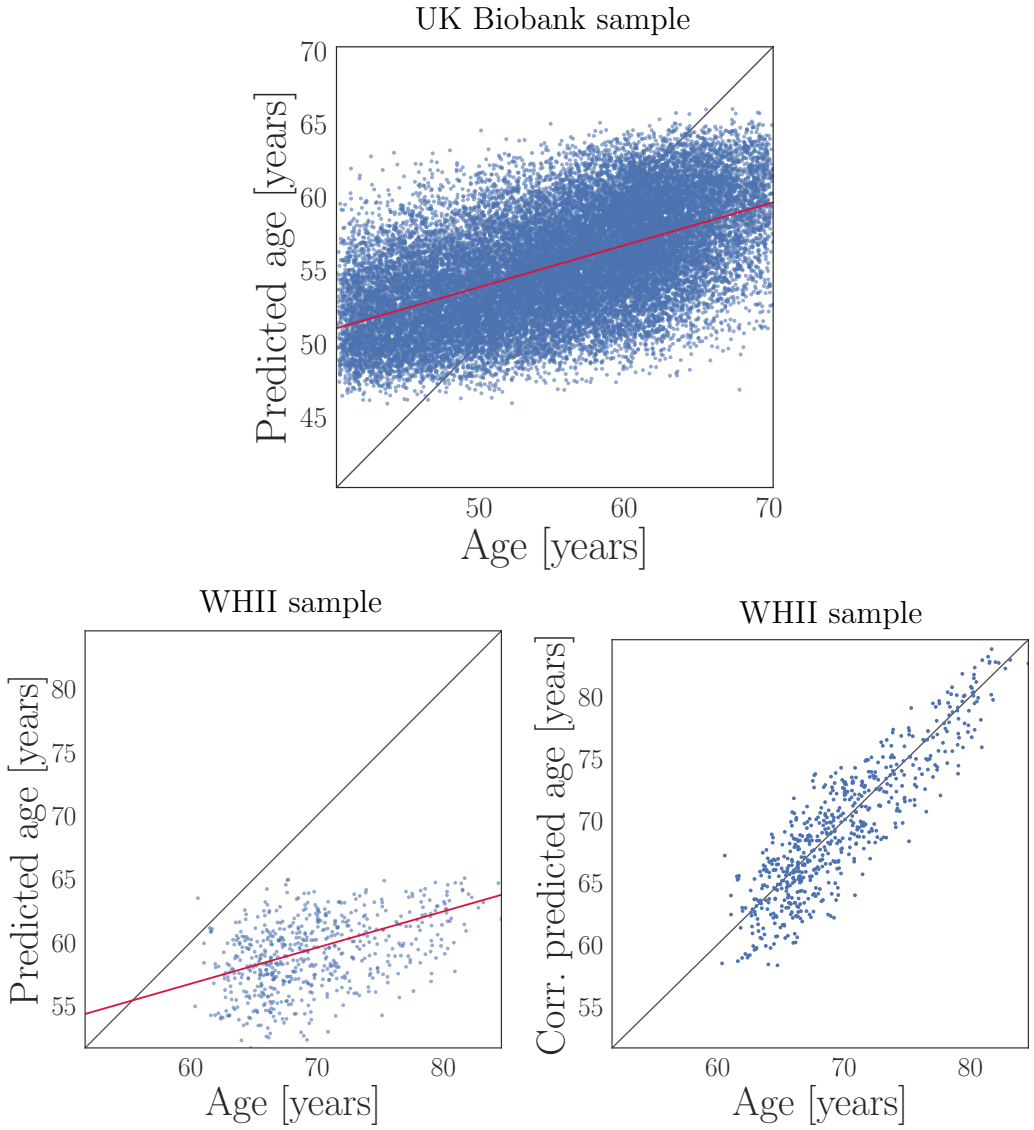
SI Figure 6: Partial dependence plots showing how the top feature from each modality affects the prediction [74]. The y-axes represent change in the prediction relative to the prediction at the leftmost value on the x-axis. For mean diffusivity (MD), which is known to increase with older age, the plot shows that MD values <0.00065 have little impact on the prediction (y-values closer to zero), but for an MD value of ≈ 0.00065 and above, the model predicts age to be higher (based on the positive change on the y-axis). However, MD values beyond 0.00069 have limited additional impact on the prediction. For volume and thickness, which are known to decrease with older age, higher values give lower age prediction. For Edge 1, higher functional connectivity values give lower age prediction. The blue shaded areas indicate confidence level. See e.g. https://www.kaggle.com/dansbecker/partial-plots for more information on partial dependence plots. AIC = Anterior Limb of Internal Capsule; Edge 1 = functional connectivity between Anterior Cingulate Gyrus and Superior Frontal Gyrus; MD = Mean Diffusivity; PHA2 is defined in SI_Glasser_2016.pdf.

Principal component analysis (PCA)

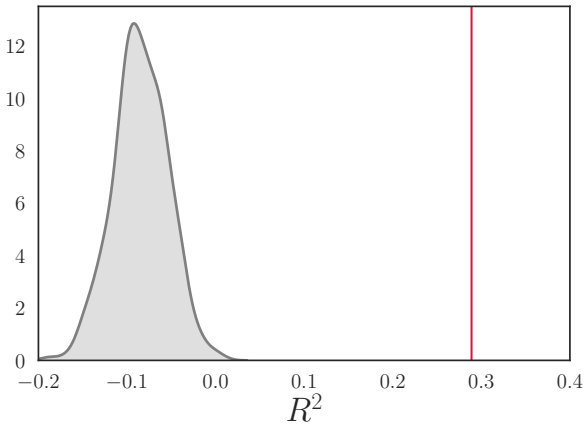
To ensure that results were consistent across dimensionality reduction methods, the models were re-run using PCA with z-transformed MRI variables. The models were run with a 70/30 train-test split and 100 bootstraps, with i) all PCA components, ii) the number of PCA components required to explain 80% of the variance in the data, and iii) the top 200 PCA components for each model as input. The results showed prediction patterns consistent with the results based on hierarchical clustering, as shown in SI Table 1.

SI Table 1: R^2 ± standard deviation for each of the brain age models run with PCA transformed features. $R^2_{\boldsymbol{NoPCA}}$ shows the R^2 from the models without any PCA transformation. $R^2_{\boldsymbol{fullPCA}}$ represents the R^2 values from the models run with all the PCA components, while $R^2_{\mathbf{80\%}}$ represents the R^2 values from the models run with the number of components required to explain 80% of the variance in the data. The number of PCA components included in the models explaining 80 % in the data are indicated in brackets. $R^2_{\mathbf{200}}$ represents the R^2 values from the models run with the top 200 PCA components as input. The % explained variance for each of the models run with the top 200 components are indicated in brackets.

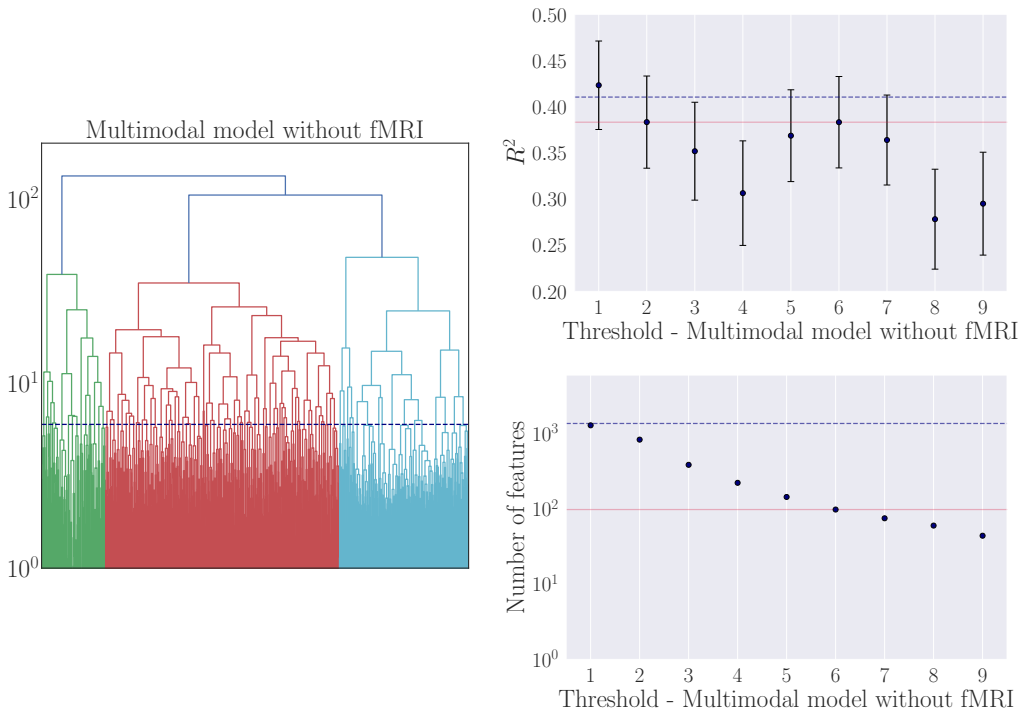
Model	$\mathbf{R^2}_{\boldsymbol{NoPCA}}$	$\mathbf{R^2}_{\boldsymbol{fullPCA}}$	$\mathbf{R^2}_{\mathbf{80\%}}$	$\mathbf{R^2}_{\mathbf{200}}$
Multimodal	0.40 ± 0.05	0.19 ± 0.06	0.21 ± 0.06 (361)	0.24 ± 0.05 (57.46%)
Gray matter	0.32 ± 0.05	0.14± 0.05	0.17 ± 0.05 (141)	0.13 ± 0.05 (87.07%)
White matter	0.31 ± 0.05	0.23 ± 0.05	0.25 ± 0.05 (23)	0.25 ± 0.05 (100%)
F. connectivity	-0.03 ± 0.04	-0.03 ± 0.03	-0.02 ± 0.04 (369)	-0.03 ± 0.04 (57.46%)



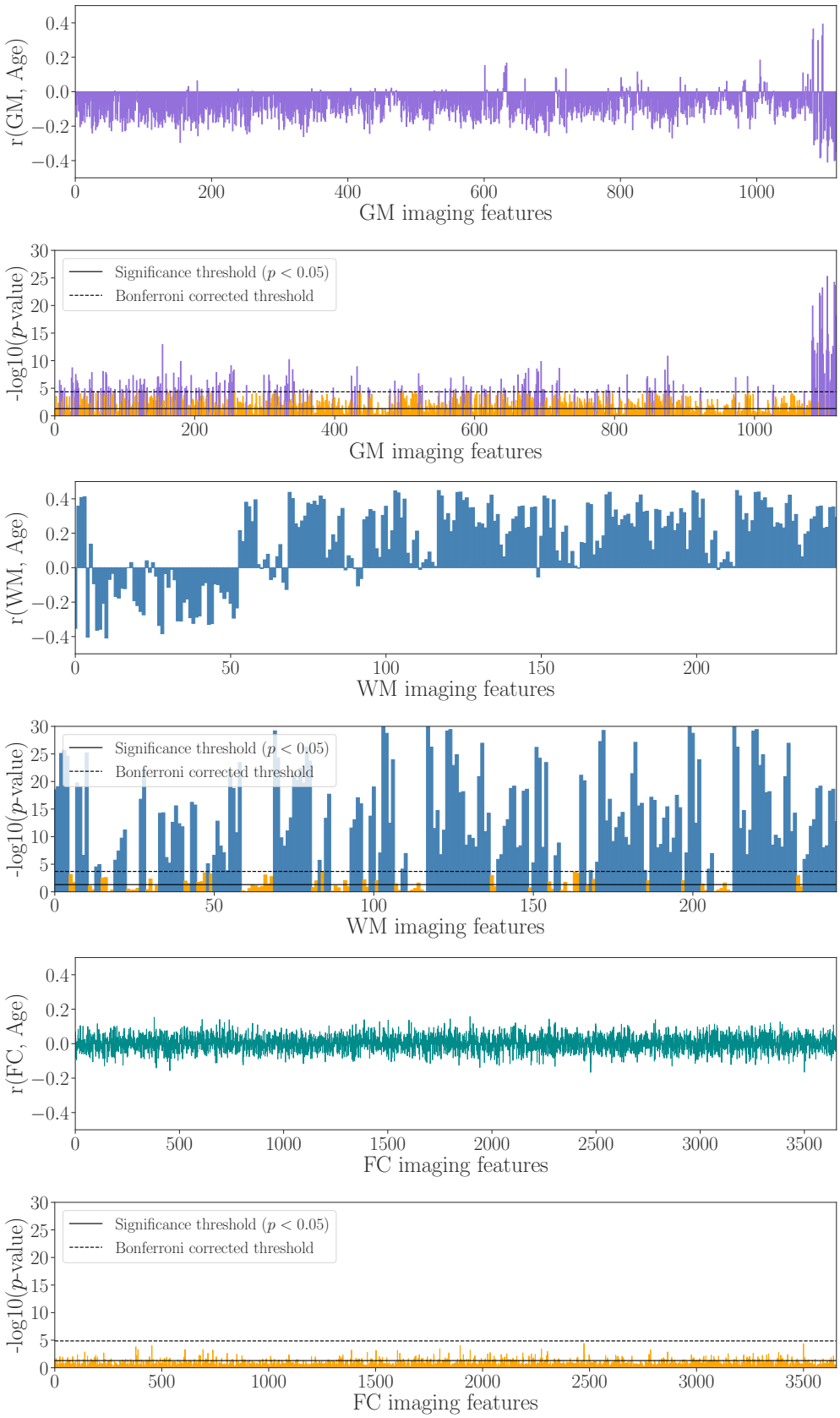
SI Figure 7: Statistical age-bias correction for the external gray matter model that was trained on the UK Biobank sample and applied to the WHII sample. Top plot: The red line shows the result of the fit that was run in the UK Biobank training set: $Y = \alpha \times \Omega + \beta$, where Y is the modelled predicted age as a function of chronological age (Ω), and α and β represent the slope and intercept. Bottom left plot: The red line shows the fit from the top plot overlaid on the WHII predictions; the WHII predictions follow the same trend as the UK Biobank predictions, but with a consistent underestimation of predicted age due to the differences in mean age between the two datasets. Bottom right plot: the WHII predictions after correcting for age-bias using the the derived values of α and β from the UK Biobank fit (top plot) with *Corrected Predicted Age* = *Predicted Age* + $[\Omega - (\alpha \times \Omega + \beta)]$.



SI Figure 8: The average R^2 ± SD for the multimodal brain age model (red vertical line) was 0.29 ± 0.10 based on the cross validation with ten splits and ten repetitions. The null distribution calculated from 1000 permutations is shown in grey, with an average R^2 ± SD of -0.08 ± 0.03. The number of permuted results from the null distribution that exceeded the mean from the cross validation was 0 (p ¡ 0.001).



SI Figure 9: Left plot: Dendrogram showing the relationships between features based on hierarchical clustering on the Spearman rank-order correlations. The y axis shows the logarithmic scale for the dissimilarity between clusters, with the horizontal blue dotted line indicating the selected threshold (4) for the multimodal model without fMRI data. Right top plot: R^2 values for each cluster threshold. The error bars indicate the standard error on the R^2 values. The blue, dotted lines indicate the R^2 for the model including all features and no threshold. The red, solid lines indicate the R^2 for the selected threshold. Right bottom plot: The y axis shows the number of features for each threshold with a logarithmic scale. The blue, dotted lines indicate the number of features for the model including all features and no threshold. The red, solid lines indicates the number of features for the selected threshold.



SI Figure 10: Correlations between every imaging feature for gray matter (GM), white matter (WM), and functional connectivity (FC) and chronological age, and the corresponding p-values. The orange features in the p-value plots represent features that are not significantly correlated to age based on the Bonferroni-corrected threshold.