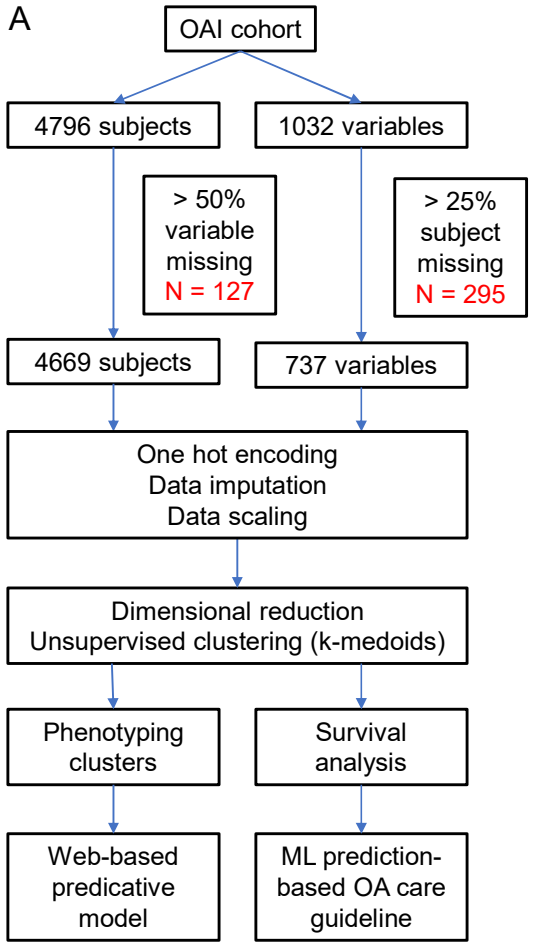
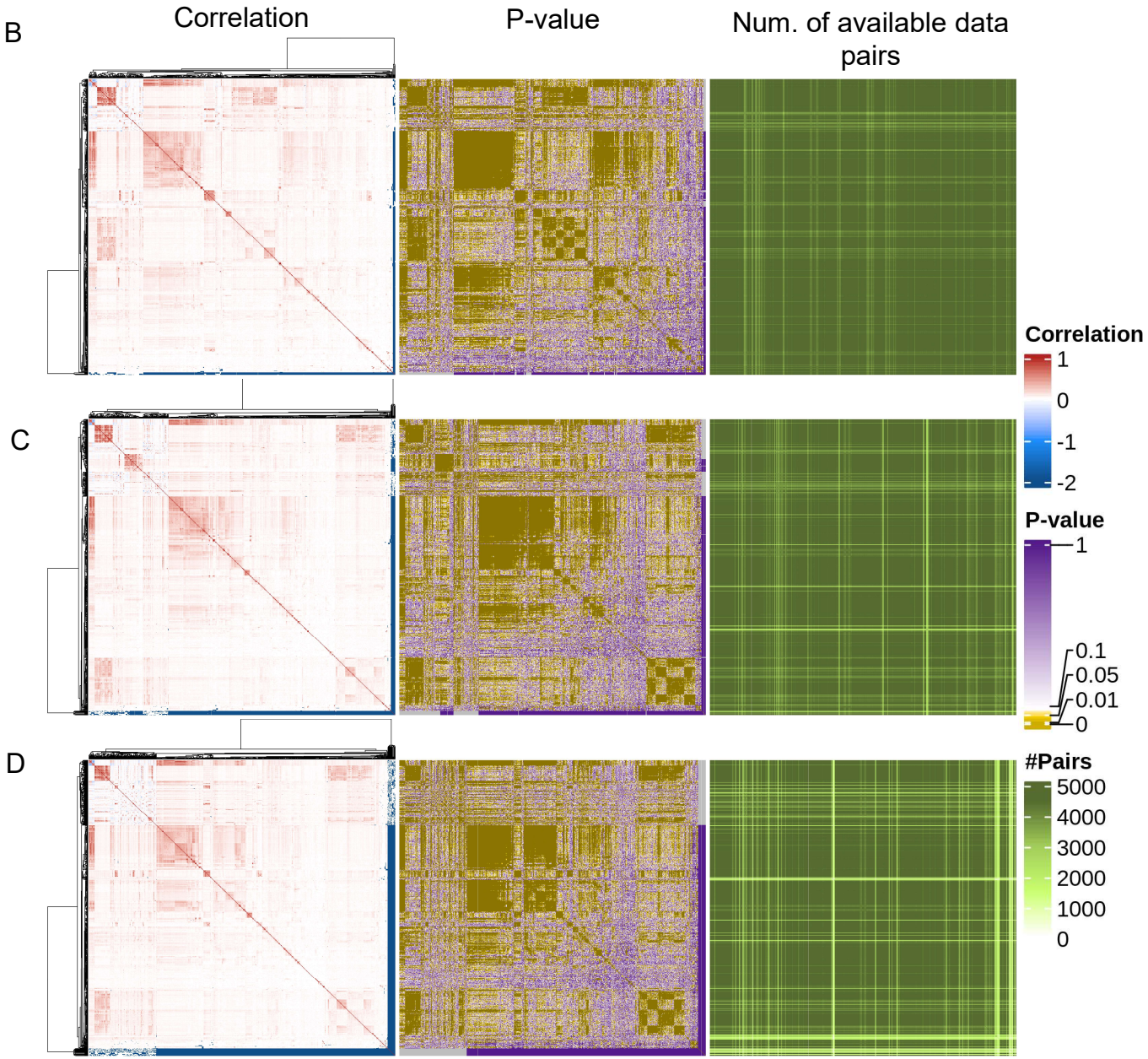
Figure S1


A
OAI cohort
4796 subjects
1032 variables
> 50% variable missing N = 127
> 25% subject missing N = 295
4669 subjects
737 variables
One hot encoding
Data imputation
Data scaling
Dimensional reduction
Unsupervised clustering (k-medoids)
Phenotyping clusters
Survival analysis
Web-based predicative model
ML prediction-based OA care guideline


B
C
D
Correlation
P-value
Num. of available data pairs
Correlation
1
0
-1
-2
P-value
1
0.1
0.05
0.01
0
#Pairs
5000
4000
3000
2000
1000
0