

Yana Hrytsenko, Yogesh Purushotham, Zhirou Li, Carmen R Isasi, Wassim Tarraf, Alberto R Ramos, Martha L Daviglus, Humberto Parada Jr., Charles DeCarli, Hector M Gonzalez, Myriam Fornage, Tamar Sofer

*Contact: tsofer@bidmc.harvard.edu, Beth Israel Deaconess Medical Center, Harvard Medical School

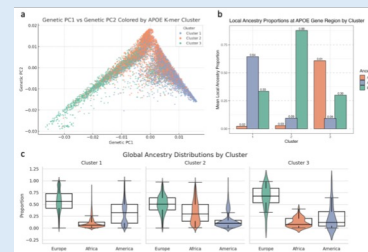
APOE, K-mer clusters and other measures of ancestry

- Focusing on the genomic region around the *APOE* gene, use short genomic sequences (k-mers) as in the figure.
- Main analysis: $k=21$.



-
- Figure 2 displays a 6x6 matrix of density scatter plots showing the relationship between six principal components (PC1 to PC6) and the APOE group. The diagonal elements are density plots for each PC. The off-diagonal elements are scatter plots of pairs of PCs, with density contours overlaid. The APOE group is color-coded: 22 (n=31) in red, 23 (n=780) in green, 24 (n=113) in blue, 33 (n=7173) in yellow, 34 (n=2118) in cyan, and 44 (n=163) in magenta.

- On the left, we visualized k-mers clusters and standard measures of ancestry: standard **genetic PCs**, **local genetic ancestry** at the **APOE gene region**, and **global genetic ancestry proportions**.
- There are some associations with all measures.



- Association analyses were adjusted for:
 - Age, years from baseline to SOL-INCA exam, sex, education level, language preference (English/Spanish), study site.

APOE variable	Estimated odds ratio	95% CI	p-value	Model AIC
Association with APOE-kmer clusters with MCI				
Cluster 2	0.92	(0.65, 1.29)	0.62	3681.59
Cluster 3	1.33	(0.96, 1.83)	0.086	
Association with APOE alleles with MCI				
APOE ε2	0.59	(0.45, 1.29)	0.53	3686.94
APOE ε4	1.00	(0.77, 1.30)	0.98	
Association with APOE-kmer clusters with significant cognitive decline				
Cluster 2	1.09	(0.88, 1.36)	0.48	6552.67
Cluster 3	1.47	(1.18, 1.84)	0.0007	
Association with APOE alleles with significant cognitive decline				
APOE ε2	1.09	(0.82, 1.45)	0.55	6569.678
APOE ε4	1.00	(0.73, 1.37)	0.97	

- **Sensitivity analyses**
 - Repeating the k-mers analysis with k=17 (weaker association), k=25 (slightly stronger).

- Repeating the k-mers analysis with **control regions**:

- *APOE* region + 4 Mega base-pairs (weaker association)
- *APOE* coordinates in chromosomes 17 and 18 (null associations)

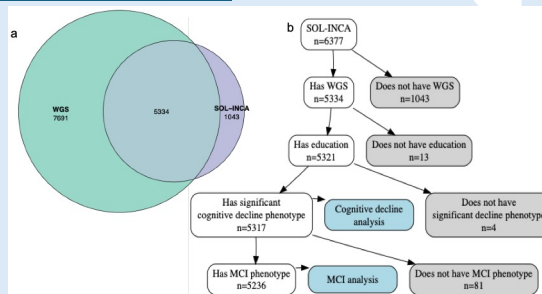
- Association analyses with k-mer PC1:
Associated with significant cognitive decline (OR=1.12; p=0.006).

- Association analyses with top k-mer:
Weak associations.

- Adjustment to global, local ancestry:
Essentially no change in associations.

- We used 13,025 individuals with Whole-Genome Sequencing (WGS) to apply the k-mers analysis and generate clusters.
- There were three resulting clusters.
- *APOE* alleles' frequencies differed by cluster.

Cluster	N	APOE ϵ 2 count	APOE ϵ 4 count	APOE ϵ 2 frequency	APOE ϵ 4 frequency
1	4496	169	987	0.038	0.220
2	5793	527	1112	0.091	0.192
3	2726	478	1089	0.175	0.398



Variable	Cluster 1	Cluster 2	Cluster 3
N (unweighted)	1770	2457	1107
Age at baseline (years)	54.58 (7.08)	55.27 (7.37)	55.61 (7.26)
Years between visit 1 and SOL-INCA visit	7.04 (1.15)	6.90 (1.16)	6.90 (1.16)
Mild Cognitive Impairment (MCI)	182 (9.9%)	246 (10.8%)	133 (13.4%)
Significant cognitive decline	517 (31.1%)	715 (29.1%)	363 (37.4%)
Sex			
Female	1157 (65.9%)	1521 (61.4%)	713 (65.8%)
Male	613 (44.1%)	936 (48.6%)	394 (44.2%)
Field center			
Bronx	249 (24.8%)	472 (20.3%)	580 (53.7%)
Chicago	605 (20.3%)	461 (8.6%)	171 (6.4%)
Miami	294 (21.4%)	912 (49.3%)	258 (29.8%)
San Diego	622 (33.5%)	612 (21.8%)	98 (10.0%)
Education			
No high school diploma or GED	798 (42.9%)	891 (32.0%)	447 (41.0%)
At most high school diploma/GED	375 (22.7%)	530 (21.2%)	250 (21.6%)
Greater than high school/GED	591 (34.6%)	1033 (46.8%)	406 (37.4%)
Hispanic/Latino heritage			
Central American	213 (8.7%)	192 (5.3%)	94 (6.6%)
Cuban	109 (10.1%)	683 (39.9%)	169 (22.3%)
Dominican	42 (2.3%)	167 (6.4%)	132 (26.8%)
Mexican	1047 (56.3%)	770 (24.8%)	130 (11.7%)
South American	153 (7.9%)	149 (3.6%)	40 (2.5%)
Puerto Rican	181 (10.0%)	440 (16.0%)	333 (25.8%)
More than one/Other heritage	21 (2.8%)	55 (4.1%)	25 (4.3%)

- We subsetting the data to SOL-INCA participants, and performed association analysis in this dataset.

Conclusions and future work

- K-mers are short genetic sequences.
- Capture added information compared to standard genetic variants.
- Correlated with population structure; independently associated with significant cognitive decline.

- Need to substantiate the evidence:
- More studies
- Identify specific sequences/haplotypes