Genetic ancestry and epigenetic age in colorectal tumors from The Cancer Genome Atlas (TCGA)

Christopher L. Benson II[1,2] , Marco Matejcic[2] , Fredrick R. Schumacher[1,3], Stephanie L. Schmit[1,2,3]

[1]Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA, [2]Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH, USA, [3]Population and Cancer Prevention Program, Case Comprehensive Cancer Center, Cleveland, OH, United States

Background

- In the United States, Colorectal cancer (CRC) is the third leading cause of cancer-related deaths in men while fourth in women.
- Among racial/ethnic populations, American Indian/Alaska Native and Black individuals have the highest rates of CRC.
- Environmental and lifestyle factors can influence methylation patterns in the genome.
- Epigenetic clocks estimate biological age from blood or tissue samples using specific CpG sites, cytosine followed by guanine from 5′ to 3′ direction, from DNA methylation (DNAm) array data.

Objective

Examine the association between global genetic ancestry proportions and epigenetic age *(i.e.,* biological age) among patients with CRC.

Methods

- DNAm array data (Illumina Human Methylation 450) was obtained from National Cancer Institute's Genomic Data Commons (GDC) on 311 subjects from The Cancer Genome Atlas (TCGA) with tumors in the colon, rectum, or rectosigmoid junction (COAD and READ).
- Beta values were generated from the DNAm data on primary tumor samples and used to estimate their epigenetic ages using two epigenetic clocks (PhenoAge and epiTOC).
- Proportions of global ancestry were estimated using ADMIXTURE based on 1000 Genomes and Population Architecture Using Genomics and Epidemiology (PAGE) as reference panels.
  - European (EUR), African (AFR), East Asian (EAS), Native American (NAT)
- Associations between global genetic ancestry and epigenetic age (outcome) for each clock were examined using linear regression, adjusted for age at diagnosis, sex, stage, and site (colon or rectum).
- Individual ancestral components (i.e. AFR, EAS, NAT) underwent transformation and were expressed in terms of additive log ratios with respect to a fixed reference component (EUR) when testing for association with genetic ancestry overall.

$$\underline{\text{Model w/o Ancestry}}$$
$$Y = \beta_0 + Covariates$$

$$\underline{\text{Model w/ Ancestry}}$$
$$Y = \beta_0 + \beta_1\left(log_2\frac{AFR}{EUR}\right) + \beta_2\left(log_2\frac{EAS}{EUR}\right) + \beta_3\left(log_2\frac{NAT}{EUR}\right) + Covariates$$

- The overall contribution of genetic ancestry to epigenetic aging was tested through a 3-degree of freedom likelihood-ratio test (3-df LRT).
- Deciles were created to group each ancestry proportion by bins (n=10) to represent a 10% increase in ancestry proportion. Each ancestry was tested for association with epigenetic age in individual linear regression models.

Table 1. Demographic and Clinical Characterisitcs of TCGA Study Participants (N=311).

	n (%)
Age At Diagnosis	
mean (SD)	64.62 (13.04)
Age Range	31 - 90
Sex (n (%))	
Male	161 (51.8)
Female	150 (48.2)
PhenoAge (in years)	
mean (SD)	94.64 (30.69)
Range	26.4 - 209.2
epiTOC (pcgtAge)	
mean (SD)	0.34 (0.10)
Range	0.12 - 0.61
Pathology Stage (n (%))	
Stage 1	50 (16.6)
Stage 2	113 (37.4)
Stage 3	93 (30.8)
Stage 4	46 (15.2)
Primary Site (n (%))	
Colon	232 (74.6)
Rectum	79 (25.4)

Epigenetic Clock Pipeline

PhenoAge (Enmix R Package)

DNAm Data (Beta Values) from tissue samples → PhenoAge Clock (513 aging related CpG Sites) → PhenoAge (Epigenetic Ages)

*PhenoAge Epigenetic Ages are expressed in years.

epiTOC (R Script)

DNAm Data (Beta Values) from tissue samples → epiTOC (385 Polycomb group target (pcgt) promotor CpG Sites) → pcgtAges (Epigenetic Ages)*

*pcgtAges are expressed in a proportion that represents a tick rate that correlates with the estimated rate of stem cell division in samples. The tick rate is defined as the increase in DNA methylation at CpG sites located in the Polycomb group target gene promoters.

Results

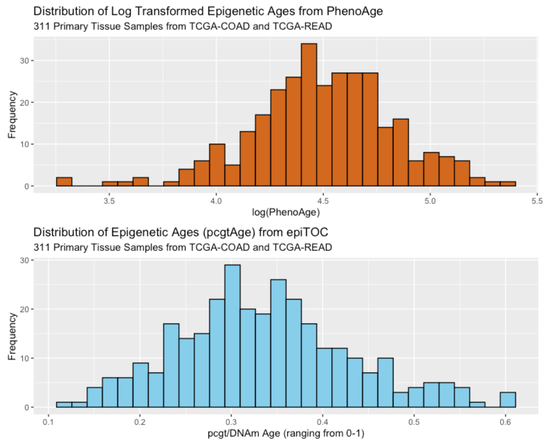


Fig 1. Epigenetic age distributions. PhenoAge epigenetic ages were transformed using a log y transformation based on Tukey's Ladder of Transformations.

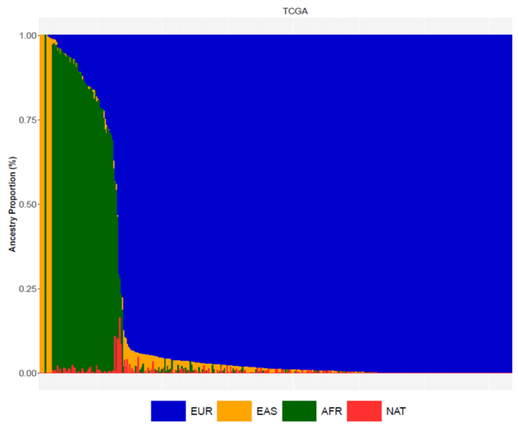
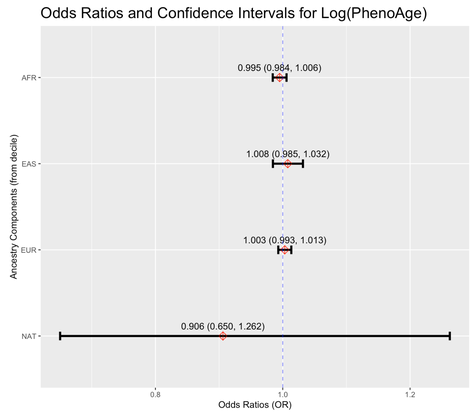


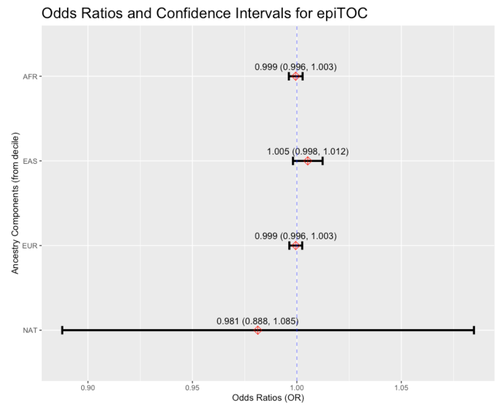
Fig 2. Overall genetic ancestry of each TCGA tumor tissue sample from ADMIXTURE using 1000 Genomes and PAGE as reference panels.

No statistically significant association was identified between epigenetic age (PhenoAge or epiTOC) and overall contribution of genetic ancestry (p > 0.05)



Ancestry	OR	Low 95% CI	High 95% CI	p-value
EUR	1.003	0.993	1.013	0.547
AFR	0.995	0.984	1.006	0.374
EAS	1.008	0.985	1.032	0.512
NAT	0.906	0.65	1.262	0.56

Fig 3. Forest Plot and table displaying OR and CI for the association between Log(PhenoAge) epigenetic ages and individual ancestry components. The proportions for each ancestry were separated into deciles.



Ancestry	OR	Low 95% CI	High 95% CI	p-value
EUR	0.999	0.996	1.003	0.741
AFR	0.999	0.996	1.003	0.746
EAS	1.005	0.998	1.012	0.149
NAT	0.981	0.888	1.085	0.712

Fig 4. Forest Plot and table displaying OR and CI for the association between epiTOC epigenetic ages and individual ancestry components. The proportions for each ancestry were separated into deciles.

Conclusions

- Preliminary analysis demonstrates that in isolation, there is no association between individual ancestry components and epigenetic age when examined using linear regression, adjusted for age at diagnosis, sex, stage, and site (colon or rectum).
- This same association was observed in models using deciles to represent a 10% increase in ancestry proportion.
- The 3-degree of freedom likelihood-ratio test (3-df LRT) found that there is no statistically significant association between the overall contribution of genetic ancestry and epigenetic aging (p-value > 0.05).

Limitations:

- We had a limited sample size for subjects with substantive non-EUR ancestral proportions.
- PhenoAge and epiTOC clocks were trained mainly on blood samples when developed.
- Global ancestry was used in this analysis rather than local ancestry.

Future Directions

- Perform a paired analysis for subjects that have DNAm data for solid normal samples.
- Replicate this analysis on paired blood samples from other studies (no DNAm data for blood in TCGA).
- Use a multi-level approach to include environmental and lifestyle factors such as neighborhood deprivation and proximity to environmental hazards and their associations with epigenetic aging.
- Include more samples with substantive non-EUR ancestorial proportions, as there is little representation of EAS and NAT ancestry in the TCGA dataset. Thus, this highlights the need for future research in more diverse populations.

References

1. Yang Z, Wong A, Kuh D, et al. Correlation of an epigenetic mitotic clock with cancer risk. Genome Biology. 2016;17(1):205. doi:10.1186/s13059-016-1064-3
2. Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. Published online 2018.
3. Teschendorff AE. A comparison of epigenetic mitotic-like clocks for cancer risk prediction. Genome Med. 2020;12(1):56. doi:10.1186/s13073-020-00752-3

Funding

This project is supported by a Cleveland Clinic Velosano 9 Pilot Award and a Burroughs Wellcome Foundation fellowship.