

Racial and Genetic Ancestry Associations with Gene Expression Patterns in a Real-World Cohort of Colorectal Cancer Patients

David Hein, MS¹; Brooke Rhead PhD²; Yannick Pouliot, PhD²; Justin Guinney, PhD²; Francisco De La Vega DSc²; and Nina Sanford, MD¹ ¹The University of Texas Southwestern Medical Center, Dallas, TX; ²Tempus Labs, Chicago, IL

Introduction

There is a growing incidence of colorectal cancer (CRC) among young adults and persistent disparities in outcomes by race/ethnicity across all ages. Gene expression signatures, as well as consensus molecular subtypes (CMS) derived from these, have been proposed to predict prognosis and therapy response in CRC. However, it is unclear whether gene expression and/or CMS are associated with racial disparities observed in CRC. We assessed whether race or genetic ancestry are associated with CMS or gene expression patterns in a deidentified cohort of 1956 CRC patients.



eparation	Data Pre								
etic ancestry s from DNA /e markers	mpus xT DNA- and RNA- eq ⁴ on tissue from colon proportion or rectum informativ								
genes, assess h effects	Generate imputed racial/ethnic categories from reported literature								
ar Subtype Ar	Consensus Molecul								
eq2 followed MMA tchEffects	patients, start from raw counts and filter rare genes removeBa								
gression on ILF	Multinomial logistic reg								
racial/ethnic c	Chi-square tests on								
t Analysis	Gene Set								
w counts and re genes	Remove MSI high and select covariates from clinical data								
VST via DES rem	TMM normalization, LIMMA-VOOM to produce precision weights								
GSVA t enrie	Multivariate linear model with empirical Bayes (LIMMA) for single gene fold change estimate								
Multivari empirical E	mROAST to test differentially expressed pathways								

Tests performed on both discrete categories and ILR transformed proportions

Contact

David Hein UT Southwestern Medical Center, Department of Radiation Oncology David.hein@utsouthwestern.edu Github: DavidHein96



previously reported to be more common among Black patients. The association of Indeterminate with Hispanic/Latino ancestry highlights the need to use diverse patient cohorts when training unsupervised learning models to improve prognosis prediction in non-White patients. Finally, our gene set analysis reveals under expression in several cancer related gene sets among NH Black patients and those with increased AFR ancestry.

				NH Black ROAST	AFR ROAST	NH Black GSVA	AFR GSVA	
		BC. RECK	- Set					
	BC. Alternative Compliment		lue For					
Hallmark Coagulation			P va					
Gene Level Log Fold Change								
					Log 2 Fo change	Log 2 Fold For Formation F		
					(-0.1 ,-0.	2] (0	.2, 0.1]	
					(-0.2, -0.3] (0.		1, 0.05]	
					(-0.3, -0.	4] (0.0	05 <i>,</i> 0.01]	
					(-0.4, -0.	5]	<0.01	
					-0.5+			
					0.6+			
			Not in gene set					

References

- 1. Beaubier N, et al. Clinical validation of the tempus xT next-generation targeted oncology sequencing assay. Oncotarget. 2019 2. Ritchie ME, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 3. Eide PW, et al. CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. Sci Rep. 2017 4. Liberzon A, et al. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011
- 5. Hänzelmann S, GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013

