TITLE: Representation of Racial Groups in Genomic Studies of Gastroenteropancreatic Neuroendocrine Neoplasms.

AUTHORS: Brendon Herring (1), Andrew Bonner (1), Catesby Mallard (1), Rachael Guenter (1), Selwyn Vickers (1), Herbert Chen (1), Clayton Yates (2), J. Bart Rose (1)

(1) University of Alabama at Birmingham, Department of Surgery

(2) Tuskegee University, Department of Biology

INTRODUCTION: Translational cancer genomics is maturing at an exponential rate, bringing invaluable findings from the bench to clinical practice. Integrated genomics will be key to discovering oncogenic pathways and therapeutic susceptibilities in gastroenteropancreatic neuroendocrine neoplasms (GEP-NENs), for which few effective therapies are available. However, not all populations are poised to benefit from these studies, as large-scale genomic analyses have consisted of mostly White/Caucasian patients. The goal of this study was to evaluate ethnoracial populations represented in genomic, transcriptomic, and epigenomic studies of GEP-NENs.

METHODS: A systematic review of the literature was conducted with structured queries to PUBMED, EMBASE, and the Cochrane Library. Abstracts were reviewed independently by two individuals for the sequencing/genetic analysis of DNA, RNA, or DNA methylation. Natural Language Processing (NLP) using the python package NLTK was then used to determine the frequency of the words “Race,” “Ethnicity,” “African American,” “Black” “Hispanic,” “Latino,” “Asian” “Native American,” “Native Hawaiian,” “Pacific Islander” “Caucasian,” and “White,” in these manuscripts. Tumor types and subject numbers by ethnoracial group were then manually evaluated. The number of subjects and the number of studies including subjects by race were evaluated with Fisher’s Exact test.

RESULTS: 313 articles conducted requisite genetic/epigenetic/transcriptomic analyses and were selected for NLP. 72 of these contained race category terms and were manually assessed for subject-level data. In total, 16/313 studies included data on the race of their subjects. 13/184 studies analyzing DNA, 4/107 analyzing RNA, and 1/54 studies analyzing Methylation included data on the race of their subjects. In the 313 studies, pancreatic NENs (pNENs) were the most common tumor type (**Fig. 1A**), followed by small intestinal NENs (siNENs) and colorectal NENs. (crNENs). In the 16 studies reporting patient race, siNENs were the most common, followed by pNENs and crNENs(**Fig. 1A**). These 16 studies included 89% White subjects (n=2032; s=16, where s = studies including White subjects), 5.8% Asian subjects (n=132, s=8), 4.0% "Other" subjects (n=93, s=11), 1.2% Black subjects (n=27, s=6), 0% Hispanic/Latino subjects, 0% Native American, and 0% Native Hawaiian/Pacific Islander subjects (**Fig. 1B**; $p < 0.01$).

CONCLUSIONS: There is little representation of racial minorities in genomic studies of GEP-NENs. Data for these populations are integral for understanding GEP-NEN biology, generalizing findings, and improving therapy.

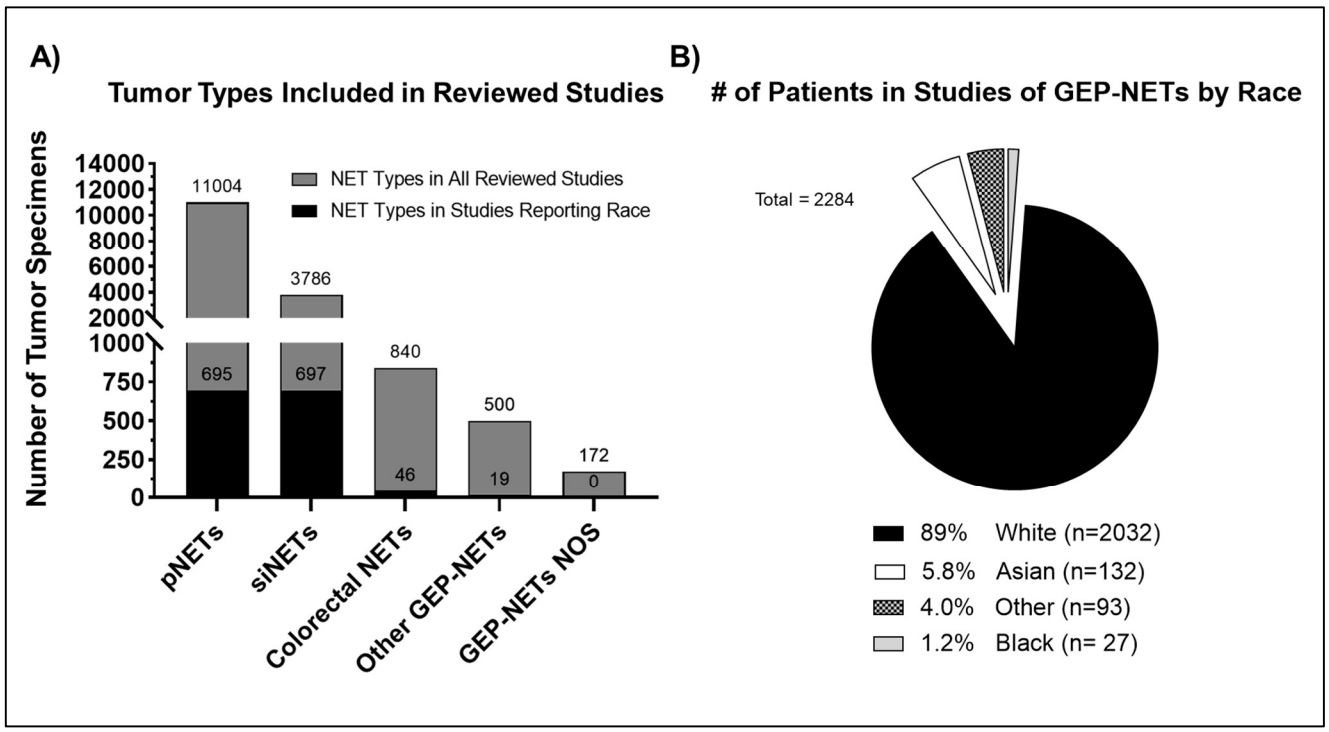

Figure 1: A) Types of GEP-NENs included in all studies that underwent NLP and studies including the race data of their subjects. B) Proportions of each race represented in studies reporting race.