



Published in final edited form as:

Cancer Discov. 2023 March 01; 13(3): 570–579. doi:10.1158/2159-8290.CD-22-0764.

Racial/ethnic and sex differences in somatic cancer gene mutations among patients with early-onset colorectal cancer

Andreana N. Holowatyj^{1,4,5}, Wanqing Wen¹, Timothy Gibbs⁶, Hannah M. Seagle^{1,5}, Samantha R. Keller^{1,5}, Digna R. Velez Edwards³, Mary K. Washington^{2,4}, Cathy Eng^{1,4}, Jose Perea^{7,8}, Wei Zheng^{1,4}, Xingyi Guo^{1,4}

¹Department of Medicine, Microbiology and Immunology

²Department of Pathology, Microbiology and Immunology

³Department of Obstetrics and Gynecology; Vanderbilt University Medical Center, Nashville, TN

⁴Vanderbilt-Ingram Cancer Center, Nashville, TN

⁵Vanderbilt University School of Medicine, Nashville, TN

⁶Meharry Medical College, Nashville, TN

⁷Molecular Medicine Unit, Department of Medicine, Biomedical Research Institute of Salamanca (IBSAL), Salamanca, Spain

⁸Department of Surgery, “Vithas Arturo Soria” University Hospital and School of Medicine, European University of Madrid, Madrid, Spain

Abstract

Molecular features underlying colorectal cancer (CRC) disparities remain uncharacterized. Here, we investigated somatic mutation patterns by race/ethnicity and sex among 5,856 non-Hispanic White (NHW), 535 non-Hispanic Black (NHB) and 512 Asian/Pacific Islander (API) CRC patients [2,016 early-onset:sequencing age<50]. NHB patients with early-onset non-hypermutated CRC, but not API patients, had higher adjusted tumor mutation rates than NHW patients. There were significant differences for *LRP1B*, *FLT4*, *FBXW7*, *RNF43*, *ATRX*, *APC* and *PIK3CA* mutation frequencies in early-onset non-hypermutated CRCs between racial/ethnic groups. Heterogeneities by race/ethnicity were observed for the effect of *APC*, *FLT4* and *FAT1* between early-onset and late-onset non-hypermutated CRC. By sex, heterogeneity was observed for the effect of *EP300*, *BRAF*, *WRN*, *KRAS*, *AXIN2* and *SMAD2*. Males and females with non-hypermutated CRC had different trends in *EP300* mutations by age group. These findings define genomic patterns of early-onset non-hypermutated CRC by race/ethnicity and sex, which yields novel biological clues into early-onset CRC disparities.

Corresponding Authors: Andreana N. Holowatyj, PhD, MS, Vanderbilt University Medical Center, 2525 West End Ave, Suite 334-G, Nashville, Tennessee 37203, Office: 615-322-0105, Fax: 615-936-8241, andreana.holowatyj@vumc.org, Xingyi Guo, PhD, Vanderbilt University Medical Center, 2525 West End Ave, Suite 330, Nashville, Tennessee 37203, Office: 615-936-3471, Fax: 615-936-8241, xingyi.guo@vumc.org.

Author Contributions: A.N.H. and X.G. participated in the conceptualization, design and interpretation of the reported experiments or results. All authors participated in the acquisition and/or analysis of data. All authors participated in the drafting and/or revising of the manuscript.

Conflict of Interest: The authors declare no conflicts of interest with this work.

Keywords

early-onset; colorectal cancer; race; ethnicity; sex; disparities; young-onset; somatic mutations; genomics; AACR Project GENIE

Introduction

The incidence of CRC among adults younger than age 50 years (early-onset CRC) has continued to steadily increase over the last several decades, leading to approximately 18,000 new cases of early-onset CRC now diagnosed annually in the United States (1). It is projected that one in ten colon cancer and one in four rectal cancer diagnoses will occur among adults younger than age 50 years by 2030 (2). Thus, to uncover the biological mechanisms that may be contributing to this alarming early-onset CRC epidemic, numerous studies have undertaken investigations into molecular features of early-onset CRC in comparison to colorectal tumors from adults diagnosed at ages 50+ years (late-onset CRC). Recent evidence from one single institution study posited that CRCs in young individuals may not biologically differ from late-onset tumors (3). However, findings from other studies worldwide are accumulating to support that early-onset CRC may harbor a distinct molecular phenotype and that tumor biology is a strong prognostic factor in early-onset CRC (4–10). Despite these key biological clues into early-onset CRC, no definitive causality has been derived for its etiology to date—which can be largely attributed to the complex and multifactorial nature of this disease.

The complexity of early-onset colorectal carcinogenesis reflects an intricate interplay of biology and genetics with health behaviors, early-life exposures, and social determinants of health. Collectively, these factors are also posited to be major drivers of pronounced disparities in the early-onset CRC burden—including by race/ethnicity as well as sex (11). However, few studies have specifically examined early-onset colorectal tumor biology *across diverse populations*, such that there remains a significant barrier in discovering potential mechanisms for the development of precision therapeutic modalities that may help mitigate a disproportionate disease burden. In the present study of 6,903 non-Hispanic White (NHW), non-Hispanic Black (NHB) and Asian or Pacific Islander (API) patients (2,016 early-onset; 29.2%) with colorectal adenocarcinoma as well as clinical-grade targeted sequencing and clinicodemographic data from the American Association for Cancer Research (AACR) Project Genomics Evidence Neoplasia Information Exchange (GENIE) international consortium (12, 13) (Table S1), we investigated tumor mutational burden (TMB) and somatic cancer gene mutation patterns of early-onset CRC by race/ethnicity and sex.

Results

TMB patterns in hypermutated colorectal tumors

In total, 9.5% of CRCs (653 of 6,903) in our cohort were hypermutated (17.78 mutations/Mb) (Figures 1A/S1A–C and Table S1). Among patients with hypermutated CRCs, young individuals (n=184) had a significantly higher TMB compared with late-onset

cases (n=469) ($P=0.008$) (Figure S2). While racial/ethnic patterns of TMB in hypermutated colorectal tumors were unable to be analyzed due to limited sample sizes, we found that males with early-onset hypermutated tumors had a higher TMB compared with females ($P=0.005$). A similar sex-specific trend was noted among patients with late-onset hypermutated colorectal tumors ($P=0.003$) (Figure S2). Given the distinct biology of tumors with hypermutation (14), these 653 cases were excluded from further study.

Distinct genomic patterns of early-onset non-hypermutated CRC

Among 6,250 non-hypermutated CRCs, tumors from patients with early-onset disease (n=1,832) had an overall lower TMB compared with late-onset non-hypermutated CRCs (n=4,418; $P=7.2\times 10^{-10}$) (Figure 1B and Figure S3). The frequency of non-silent somatic mutations for the most commonly mutated genes in early-onset and late-onset non-hypermutated CRCs is presented in Figure S4. Overall, young patients with non-hypermutated colorectal tumors had significantly higher odds of presenting with non-silent mutations in *TP53*, *LRP1B*, *TCF7L2*, and *FBXW7* versus late-onset cases after adjustment for sex, race/ethnicity, tumor site and histology, sequencing assay, sample type and TMB (Figure 1C and Table S2). In contrast, young patients had decreased odds of presenting with *KDR*, *FLT4* and *AMER1* non-silent mutations in non-hypermutated tumors. Notably, patterns for *TP53*, *LRP1B*, and *TCF7L2* persisted after adjusting for multiple comparisons (all FDR<0.05). Together, these results suggest that young patients with non-hypermutated CRCs harbor unique non-silent somatic mutation patterns compared with late-onset cases—providing additional evidence to support distinct tumor characteristics of early-onset non-hypermutated CRC in a large and diverse patient cohort.

Racial/ethnic differences in somatic cancer gene mutations among patients with early-onset non-hypermutated CRC

Given the significant variation in early-onset CRC incidence and outcomes across racial and ethnic groups (15, 16), we next sought to explore genomic patterns of early-onset CRC between NHW, NHB and API individuals. Among young patients with non-hypermutated colorectal tumors, NHB patients (n=157)—but not API patients (n=164)—had a significantly higher TMB compared with NHW patients (n=1,511) ($P_{NHB/NHW}=0.01$) (Figure 2A). A similar trend was also observed between NHB and NHW individuals with late-onset non-hypermutated CRC ($P_{NHB/NHW}=0.009$) (Figure 2B). Together, the higher TMB observed in non-hypermutated colorectal tumors of NHB individuals yields potential clinical relevance given the emerging role of TMB as a predictive biomarker for therapeutic response.

Among API patients with non-hypermutated CRCs (n=469), early-onset non- cases had decreased odds of presenting with non-silent mutations in *APC*, *RIF1* and *PIK3CA* compared with API individuals ages 50+ years at cancer sequencing in adjusted models (Figure 2C and Tables 1/S3). In contrast, young APIs had increased odds of presenting with non-silent mutations in *FAT1*, *FLT4*, *FBXW7* and *MTOR* versus API individuals with non-hypermutated late-onset CRC. Young NHB patients with non-hypermutated CRCs (n=157) had statistically significantly higher odds of presenting with *ATRX* non-silent mutations versus late-onset cases in adjusted models (Figure 2D and Tables 1/S3). While the results

observed across NHB and API populations did not persist after adjustment for multiple testing, this may be in part attributable to limited sample size and warrants additional and independent study in diverse cohorts. Within the NHW population, individuals with early-onset non-hypermutated CRCs had higher odds of presenting with *LRP1B*, *TP53*, *TCF7L2*, *SMAD3* and *FBXW7* non-silent mutations (all $P < 0.02$) and decreased odds of presenting with *KDR*, *FLT4*, *RNF43* and *BRAF* mutations (all $P < 0.04$) versus late-onset non-hypermutated CRC cases in adjusted models (Figure 2E and Tables 1/S3). We also observed that the patterns for *TP53*, *LRP1B*, and *TCF7L2* in non-hypermutated colorectal tumors from early-onset cases within the NHW population persisted after adjusting for multiple comparisons (*TP53*: Odds ratio [OR] 1.39, 95% CI 1.20–1.61, $P = 1.36 \times 10^{-5}$, FDR=0.001; *LRP1B*: OR 4.75, 95% CI 2.21–10.23, $P = 6.77 \times 10^{-5}$, FDR=0.003; *TCF7L2*: OR 1.45, 95% CI 1.17–1.80, $P = 0.0008$, FDR=0.02) (Tables 1/S3).

Of these fifteen identified genes, significant heterogeneities across racial/ethnic groups were observed for the effect of *APC*, *FLT4* and *FAT1* between early-onset and late-onset non-hypermutated CRC cases (Cochran's Q-test: $P_{het} = 0.005$, 0.01, and 0.048, respectively) (Table 1). Moreover, statistically significantly different mutation frequencies in non-hypermutated CRCs among young patients across racial/ethnic groups were observed for seven genes, including: *FLT4* (Chi-square test: $P = 0.0006$), *FBXW7* ($P = 0.04$), *RNF43* ($P = 0.046$), *LRP1B* ($P = 2.45 \times 10^{-10}$), *APC* ($P < 0.0001$), *PIK3CA* ($P = 0.003$) and *ATRX* ($P = 0.03$) (Table 1). In summary, these findings point to unique somatic gene mutation landscapes by race/ethnicity *specifically within* the population of individuals with early-onset non-hypermutated CRC.

Sex differences in non-silent somatic gene mutation profiles of early-onset non-hypermutated CRC

The biological features contributing to early-onset CRC disparities by sex remain presently unknown (11). Therefore, we also sought to examine sex-specific differences in non-silent somatic gene mutation profiles of early-onset non-hypermutated CRC cases in our cohort. Investigation of TMB among 1,832 early-onset non-hypermutated CRC cases by sex revealed that males presented with a lower TMB versus females after adjusting for race/ethnicity, tumor site and histology, sequencing assay and sample type, although this difference did not reach statistical significance ($P = 0.07$) (Figure 3A). A significant pattern was observed for TMB by sex among 4,418 patients with late-onset non-hypermutated CRC in adjusted models ($P = 0.004$) (Figure 3A).

Among females, young patients with non-hypermutated CRC had statistically significantly lower odds of presenting with non-silent mutations in *EP300*, *AXIN2*, *WRN*, *BRAF* and *KDR* compared with late-onset cases in adjusted models. In contrast, females with early-onset non-hypermutated CRC had statistically significantly higher odds of presenting with *TP53*, *SMAD2*, *APC*, *TCF7L2* and *LRP1B* non-silent mutations in adjusted models (Figure 3B and Table S4). In particular, our observation that young female patients with non-hypermutated CRC had 54% increased odds of presenting with a *TP53* mutation persisted after adjustment for multiple comparisons (OR 1.54, 95% CI 1.26–1.88, $P = 2.73 \times 10^{-5}$,

FDR=0.002) (Table S4). Associations for *BRAF* and *EP300* reached marginal significance after FDR adjustment (FDR=0.057 and 0.09, respectively).

Young males with non-hypermuted CRC were statistically significantly more likely to present with non-silent mutations in *TCF7L2* and *TP53*, and less likely to present with *KRAS* mutations, versus males with late-onset non-hypermuted CRC (Figure 3C and Table S4). Although these findings did not remain significant after adjustment for multiple comparisons, the patterns for *TCF7L2* and *TP53* among young males are consistent with our observations among young females. In contrast to the observation that females with early-onset non-hypermuted CRC were 70% less likely to present with non-silent somatic mutations in *EP300* (females: OR 0.30, 95% CI 0.13–0.67, $P=0.004$; FDR=0.09), young males were 59% more likely to present with *EP300* non-silent mutations (OR 1.59, 95% CI 1.04–2.43, $P=0.03$; FDR=0.64) (Figure 3D and Table S4).

Of these eleven identified genes, significant heterogeneities between males and females were observed for the effect of *EP300*, *BRAF*, *WRN*, *KRAS*, *AXIN2* and *SMAD2* between early-onset and late-onset non-hypermuted CRC cases in our cohort (Cochran's Q-test: $P_{het}=0.0004$, 0.002, 0.03, 0.03, 0.04, and 0.04; respectively) (Table S4). Moreover, unique *EP300* mutation frequencies were observed in non-hypermuted CRCs among young patients by sex (Chi-square test: $P=0.00003$). Differences in *KRAS*, *AXIN2*, *WRN*, *BRAF* and *LRP1B* mutation rates by sex were also noted in non-hypermuted colorectal tumors of young patients (all $P<0.03$) (Figure 3B–D, Table S4). Together, these results indicate that differences in the molecular landscape of early-onset non-hypermuted CRC persist between males and females and may present potential targets for future validation and mechanistic studies among early-onset non-hypermuted CRC patients by sex.

Discussion

Here, we defined distinct molecular patterns of early-onset non-hypermuted CRC by race/ethnicity and sex using clinical-grade sequencing of colorectal adenocarcinoma. This work was performed using a large cohort of 6,903 NHW, NHB and API individuals from the AACR Project GENIE international consortium. Specific to non-hypermuted colorectal tumors, we observed striking differences for *APC* mutation rates in young patients across racial/ethnic groups—as 59% of API individuals, 76% of NHB individuals and 75% of NHW individuals with early-onset CRC harbored a non-silent *APC* mutation in non-hypermuted tumors. We also found significant heterogeneity for the effect of *APC* between early-onset and late-onset cases by race/ethnicity in non-hypermuted tumors. Indeed, our observations for *APC* align with prior publications (17, 18), including a recent study of early-onset CRC patients—including 137 non-Hispanic Asian, 128 NHB and 105 White Hispanic individuals, that explored tumor mutation patterns for 22 genes. Among early-onset CRC patients with *both* non-hypermuted and hypermutated tumors, Hein and colleagues noted differences in *APC* mutation rates when comparing racial/ethnic groups—with lower rates of *APC* mutations observed among non-Hispanic Asian patients compared with NHW patients (17). This study also revealed that 21% of tumors from young non-Hispanic Asian individuals had *FBXW7* mutations versus 15% of young NHW or 11.7% of young NHB patients. Herein, mutation rates for *FBXW7* (17% of young API, 11%

NHW, and 10% NHB patients), as well as for *FLT4*, *RNF43*, *LRP1B*, *PIK3CA* and *ATRX*, also significantly differed across racial/ethnic groups for early-onset non-hypermutated tumors. One advantage to our approach was that we investigated approximately 4-fold more genes and restricted all analyses to non-silent somatic gene mutations specific to patients with non-hypermutated colorectal tumors. Moreover, our comparison of early-onset and late-onset cases provided the opportunity to identify genomic features distinct to early-onset non-hypermutated CRC across both race/ethnicity and sex.

While these findings provide novel clues into biological mechanisms that may be underpinning the disproportionate early-onset CRC burden across racial/ethnic groups, one important acknowledgement with respect to these findings is that race and ethnicity is a social construct. It is vital to consider the role of several complex and related factors, particularly genetic ancestry as a biological construct, in early-onset CRC disparities (19) and in the interpretation of our results. It is equally important to compare our present findings with recent work that explored genomic features of CRC by genetic ancestry (8). Among 33,770 individuals of European ancestry and 5,301 individuals of African ancestry diagnosed with CRC—including hypermutated tumors, Myer et al. (8) observed that among all samples and specific to microsatellite stable, *POLE/POLD-1* negative colorectal cancers (TMB<10), the median TMB was significantly higher among individuals of African vs European ancestry. While more than half of all cases in that cohort were ages 59 years and older at cancer sequencing, in our present study we noted similar TMB patterns by race/ethnicity for both early-onset and late-onset non-hypermutated CRCs. In particular, we found that for young patients with non-hypermutated colorectal tumors, NHB patients—but not API patients—had a significantly higher TMB compared with NHW patients. As genomic patterns of CRC have not yet been explored among individuals of East/South Asian ancestry to date, this also suggests that independent validation of our results in future cohorts with diverse, well-annotated early-onset CRC cases and available genetic ancestry data will be vital to accelerate the translation of these distinct early-onset CRC patterns by race/ethnicity into clinical application and reduce marked disparities in this disease burden.

Beyond these racial/ethnic differences in somatic cancer gene mutations for early-onset non-hypermutated CRCs, our study is the first to our knowledge to identify genomic patterns specific to early-onset CRC by sex. As a biological variable, sex affects the function of the immune system. The complex interplay and effects of genetics, hormones and the environment (e.g. gut microbiome) can all contribute to sex differences in immune responses (20) and to variations in the burden of CRC. This is supported by a comprehensive untargeted metabolomics study of colon tumor and normal tissues from patients age 55+ years, where Cai and colleagues revealed sex-specific metabolic subphenotypes in colon cancer (21). Here, we showed sex-specific differences for *EP300*, *KRAS*, *AXIN2*, *WRN*, *BRAF* and *LRP1B* mutation rates specific to non-hypermutated colorectal tumors of young patients. We also identified significant heterogeneity by sex for the effect of *EP300*, *BRAF*, *WRN*, *KRAS*, *AXIN2* and *SMAD2* between early-onset and late-onset non-hypermutated CRC cases. Together, these findings support a hypothesis that differences in tumor biology may be contributing to a sex-specific disease burden in early-onset CRC for further study. As males harbor a 12% to 18% increased hazard of disease-specific death compared with females after early-onset CRC diagnosis (15), further investigation into the

prognostic significance of these genes in early-onset CRC by sex may also yield significant implications in the clinical setting.

Use of clinical-grade targeted sequencing and clinicodemographic data for nearly 7,000 pathologically-confirmed colorectal adenocarcinoma cases—of which nearly 30% had early-onset disease—from AACR Project GENIE (12, 13) is a considerable strength of this present study. It is also of value to draw on limitations of this work. As the present study was limited to individuals who self-identified as NHW, NHB and API, we do not yet know how diversity within these groups (e.g. Asian *versus* Pacific Islander individuals) or in other populations (e.g. Hispanic, American Indian populations) contribute to the biology of early-onset CRC disparities. GENIE data largely stems from tertiary care centers and may not completely represent the target populations. Although the genomic data provided by GENIE is clinical-grade and passed through stringent processing pipelines prior to release (22), variant calling was not independently validated by orthogonal approaches. The data released also precluded our ability to derive genetic ancestry or reliably define microsatellite instability (MSI) status for CRC cases. Further, GENIE does not make available information on clinical outcomes which limited our ability to investigate the possible prognostic value of these genes in early-onset non-hypermutated CRC by race/ethnicity or sex. GENIE also does not presently release tumor stage, grade, primary colon tumor site codes, or treatment history data. Consequently, we adjusted for tumor site (colon *vs* rectum), histology and primary sample type in our study. To consider that primary sample type (primary tumor versus metastasis) may contribute to somatic mutation differences by race/ethnicity and sex in early-onset non-hypermutated CRC, we also repeated our primary TMB analyses while excluding patients with metastatic tissue used for clinical-grade sequencing with concordant results. However, it is still possible that molecular patterns observed herein may be in part related to disease stage and warrant validation in independent cohort studies. While we were able to exclude 653 patients with hypermutated tumors to specifically focus on non-hypermutated CRC, information on family history of cancer, as well as germline genetic features, were also unavailable for query.

In conclusion, this study provides first-of-its-kind evidence to our knowledge that molecular features of early-onset CRC may differ by both race/ethnicity and sex. We also defined sex and racial/ethnic-specific differences in tumor mutational burden among early-onset CRC patients that may begin to better inform treatment decision-making. Together, these findings warrant subsequent epidemiologic and laboratory-based studies for validation from which the knowledge gained could yield unprecedented mechanistic insights into the biology underlying early-onset CRC disparities.

Methods

Study population

Next-generation clinical sequencing data from tumor tissues and associated pathology reports have been released by the AACR Project GENIE consortium (12, 13). This study has been granted data access through Database of Genotypes and Phenotypes (dbGap) project #24541. Somatic cancer gene mutation data as well as clinicopathologic and demographic data for CRC cases were downloaded from the GENIE project via Synapse (release 11.0)

(<http://www.synapse.org/genie>). This study was exempt from IRB approval and informed consent, as de-identified GENIE data are publicly available (12, 13, 22). A total of 6,903 pathologically-confirmed colorectal adenocarcinoma cases, with a unique patient record and matched clinical and sequencing data who self-identified as NHB, NHW, or API, were included in the present study.

Available clinical and pathologic data for CRC cases with clinical-grade targeted sequencing in GENIE included: site (colon, rectum), histology (colon, rectal and colorectal adenocarcinoma; colorectal mucinous adenocarcinoma), and sample type (primary tumor or metastatic site). Demographic data included: sex, race and ethnicity (NHW, NHB, API), age at sequencing—a surrogate for diagnosis age (23), and sequencing center and assay (panel/platform).

Clinical-grade targeted sequencing data

Somatic mutation data from tumor tissues have been previously generated using clinical-grade targeted sequencing panels from multiple sequencing centers (12, 13). Detailed summaries of sequencing pipelines—including distributions of library selection, library strategy, platform, and specimen tumor cellularity; coverage and alteration types per panel/pipeline; preservation techniques; sequence assay genomic information; and genomic profiling at each center; have been previously described and are publicly accessible in the AACR GENIE Data Guide (22). Read depth for CRC tissues across the 10 sequencing centers included in this study are listed in Table S5.

The bioinformatics pipelines used to detect mutations are also described in-depth in the AACR GENIE Data Guide (22), including: data pre-processing and alignment of reads, quality filters/controls, single nucleotide somatic mutation and small insertion and deletion (indel) calls, and filtering of putative germline single nucleotide variants and indels. GENIE has applied a stringent filtering pipeline to remove putative germline variants and minimize artifacts (e.g. using pooled blood samples as controls, existing databases of known artifacts, and common germline variants from the 1000 Genomes Project or Exome Sequencing Project with allele frequencies >0.1%) to ensure consistent calling of somatic variations in tumor tissues, as well as to minimize artifacts and germline events. GENIE has provided extensive functional annotation for somatic mutations based on curated bioinformatics analysis of functional genomic databases. To focus on putative functional mutations, we limited our analyses to non-silent mutations (e.g. bin variable for mutation carrier vs non-carrier), which includes missense, splicing, nonsense, truncating, frameshift insertion and deletions, and non-frameshift deletions.

Tumor mutational burden and hypermutation status

We analyzed sequencing panel coverage for each sequencing assay (panel/platform) based on relevant genomic information released by Project GENIE (Synapse) (<http://www.synapse.org/genie>) with detailed covered gene regions for each sequencing assay (22). We calculated the total covered genomic regions based on the intragenic regions included in panels for each sequencing assay. Patients with sequencing assay coverage of less than 500kb target regions were excluded from our study. TMB for each CRC case was quantified

by the total number of somatic mutations per 1Mb in tumor tissue. To focus our analyses to non-hypermutated CRC cases, a total of 653 cases with 17.78 somatic mutations/1Mb (defined as hypermutated CRC) were removed based on our conserved inflection point estimation of the TMB distribution (Figure 1A). Non-hypermutated CRCs were defined as tumors with fewer than 17.78 somatic mutations/1Mb.

Statistical analysis

Clinical and demographic features of the study population were summarized by frequency. Given the observed variation in TMB across individual sequencing assays (platforms/panels) (Figure S3), all analyses were adjusted for sequencing assay in our study. Comparison of TMB between groups (early-onset vs late-onset; sex; and race/ethnicity) was evaluated using multivariable linear regression adjusted for patient sex, race/ethnicity, colorectal tumor site and histology, sequencing assay, and sample type, as appropriate. Consequently, the residual of adjusted mutation rates was presented as a proxy to visualize TMB using multivariable linear regression models adjusted for patient sex, race/ethnicity, colorectal tumor site and histology, sequencing assay, and sample type, as appropriate (Figures 1B, 2A–B, 3A and S2).

The baseline mutation probability for each gene was estimated based on non-silent somatic mutation frequency calculated from mutation carriers divided by total cases as we have previously described (24). Overall comparison of non-silent somatic mutations between early-onset and late-onset non-hypermutated CRC cases, as well as comparisons by racial/ethnic groups and sex, was performed using multivariable logistic regression analyses adjusted for patient sex, race/ethnicity, colorectal tumor site and histology, sequencing assay, sample type and TMB, as appropriate. All co-variables were used as fixed effects. To control for multiple comparisons, false discovery rate (FDR) correction was performed on the nominal *P*-values derived from our association analyses.

Mutation frequencies for early-onset versus late-onset non-hypermutated CRC cases were visualized using bar graphs. Differences in mutation frequencies for each gene of interest by race/ethnicity and sex for early-onset non-hypermutated CRC cases were compared using Chi-square tests. Heterogeneity tests were conducted using the Cochran's Q-test. All statistical tests were two-sided, with $P < 0.05$ considered to be statistically significant. Analyses were conducted using R software version 3.3.3 (R Project for Statistical Computing).

Data availability statement

Data for AACR Project GENIE are available at <http://www.synapse.org/genie>, with terms of access provided at https://www.aacr.org/wp-content/uploads/2022/03/GENIE_data_guide_11.0-public.pdf. Data supporting the findings from this study are also available from the corresponding authors upon reasonable request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We acknowledge all the families and clinicians who contributed to the American Association of Cancer Research (AACR)'s Project Genomics Evidence Neoplasia Information Exchange (GENIE) international clinicogenomic data-sharing consortium.

Research Support:

A.N.H. was supported by the National Institutes of Health K12 HD043483 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. This work was also supported by National Institutes of Health/National Cancer Institute grants: R37 CA227130 (X.G.), R01 CA188214 (W.Z.); and by the American Cancer Society #IRG-19-139-59 (A.N.H.).

References

1. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(3):145–64. [PubMed: 32133645]
2. Bailey CE, Hu CY, You YN, Bednarski BK, Rodriguez-Bigas MA, Skibber JM, et al. Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975–2010. *JAMA Surg.* 2015;150(1):17–22. [PubMed: 25372703]
3. Cercek A, Chatila WK, Yaeger R, Walch H, Fernandes GDS, Krishnan A, et al. A Comprehensive Comparison of Early-Onset and Average-Onset Colorectal Cancers. *J Natl Cancer Inst.* 2021;113(12):1683–92. [PubMed: 34405229]
4. Holowatyj AN, Gigic B, Herpel E, Scalbert A, Schneider M, Ulrich CM, et al. Distinct Molecular Phenotype of Sporadic Colorectal Cancers Among Young Patients Based on Multiomics Analysis. *Gastroenterology.* 2020;158(4):1155–8 e2. [PubMed: 31730769]
5. Lieu CH, Golemis EA, Serebriiskii IG, Newberg J, Hemmerich A, Connelly C, et al. Comprehensive Genomic Landscapes in Early and Later Onset Colorectal Cancer. *Clin Cancer Res.* 2019;25(19):5852–8. [PubMed: 31243121]
6. Kirzin S, Marisa L, Guimbaud R, De Reynies A, Legrain M, Laurent-Puig P, et al. Sporadic early-onset colorectal cancer is a specific sub-type of cancer: a morphological, molecular and genetics study. *PLoS One.* 2014;9(8):e103159.
7. Willauer AN, Liu Y, Pereira AAL, Lam M, Morris JS, Raghav KPS, et al. Clinical and molecular characterization of early-onset colorectal cancer. *Cancer.* 2019;125(12):2002–10. [PubMed: 30854646]
8. Myer PA, Lee JK, Madison RW, Pradhan K, Newberg JY, Isasi CR, et al. The Genomics of Colorectal Cancer in Populations with African and European Ancestry. *Cancer Discovery.* 2022;12(5):1282–93. [PubMed: 35176763]
9. Jin Z, Dixon JG, Fiskum JM, Parekh HD, Sinicrope FA, Yothers G, et al. Clinicopathological and Molecular Characteristics of Early-Onset Stage III Colon Adenocarcinoma: An Analysis of the ACCENT Database. *JNCI: Journal of the National Cancer Institute.* 2021;113(12):1693–704. [PubMed: 34405233]
10. Eng C, Jacome AA, Agarwal R, Hayat MH, Byndloss MX, Holowatyj AN, et al. A comprehensive framework for early-onset colorectal cancer research. *Lancet Oncol.* 2022;23(3):e116–e28. [PubMed: 35090673]
11. Holowatyj AN, Perea J, Lieu CH. Gut instinct: a call to study the biology of early-onset colorectal cancer disparities. *Nat Rev Cancer.* 2021;21(6):339–40. [PubMed: 33833408]
12. AACR Project GENIE Consortium. AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov.* 2017;7(8):818–31. [PubMed: 28572459]
13. Pugh TJ, Bell JL, Bruce JP, Doherty GJ, Galvin M, Green MF, et al. AACR Project GENIE: 100,000 Cases and Beyond. *Cancer Discovery.* 2022;12(9):2044–57. [PubMed: 35819403]
14. Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, et al. Comprehensive Analysis of Hypermutation in Human Cancer. *Cell.* 2017;171(5):1042–56.e10. [PubMed: 29056344]

15. Holowatyj AN, Ruterbusch JJ, Rozek LS, Cote ML, Stoffel EM. Racial/Ethnic Disparities in Survival Among Patients With Young-Onset Colorectal Cancer. *J Clin Oncol*. 2016;34(18):2148–56. [PubMed: 27138583]
16. Theuer CP, Wagner JL, Taylor TH, Brewster WR, Tran D, McLaren CE, et al. Racial and ethnic colorectal cancer patterns affect the cost-effectiveness of colorectal cancer screening in the United States. *Gastroenterology*. 2001;120(4):848–56. [PubMed: 11231939]
17. Hein DM, Deng W, Bleile M, Kazmi SA, Rhead B, De La Vega FM, et al. Racial and Ethnic Differences in Genomic Profiling of Early Onset Colorectal Cancer. *J Natl Cancer Inst*. 2022;114(5):775–8. [PubMed: 35134211]
18. Liu Z, Yang C, Li X, Luo W, Roy B, Xiong T, et al. The landscape of somatic mutation in sporadic Chinese colorectal cancer. *Oncotarget*. 2018;9(44):27412–22. [PubMed: 29937994]
19. Eng C, Holowatyj AN. Colorectal Cancer Genomics by Genetic Ancestry. *Cancer Discov*. 2022;12(5):1187–8. [PubMed: 35491647]
20. Klein SL, Flanagan KL. Sex differences in immune responses. *Nature Reviews Immunology*. 2016;16(10):626–38.
21. Cai Y, Rattray NJW, Zhang Q, Mironova V, Santos-Neto A, Hsu K-S, et al. Sex Differences in Colon Cancer Metabolism Reveal A Novel Subphenotype. *Scientific Reports*. 2020;10(1):4905. [PubMed: 32184446]
22. American Association for Cancer Research (AACR) Project GENIE. [cited 2022 Jul 1]. Available from: https://www.aacr.org/wp-content/uploads/2022/03/GENIE_data_guide_11.0-public.pdf.
23. Gagan J, Van Allen EM. Next-generation sequencing to guide cancer therapy. *Genome Medicine*. 2015;7(1):80. [PubMed: 26221189]
24. Chen Z, Wen W, Beeghly-Fadiel A, Shu XO, Díez-Obrero V, Long J, et al. Identifying Putative Susceptibility Genes and Evaluating Their Associations with Somatic Mutations in Human Cancers. *Am J Hum Genet*. 2019;105(3):477–92. [PubMed: 31402092]

Statement of Significance

NHBs, but not APIs, with early-onset non-hypermuted CRC had higher adjusted tumor mutation rates versus NHWs. Differences for *FLT4*, *FBXW7*, *RNF43*, *LRP1B*, *APC*, *PIK3CA* and *ATRX* mutation rates between racial/ethnic groups and *EP300*, *KRAS*, *AXIN2*, *WRN*, *BRAF* and *LRP1B* mutation rates by sex were observed in tumors of young patients.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

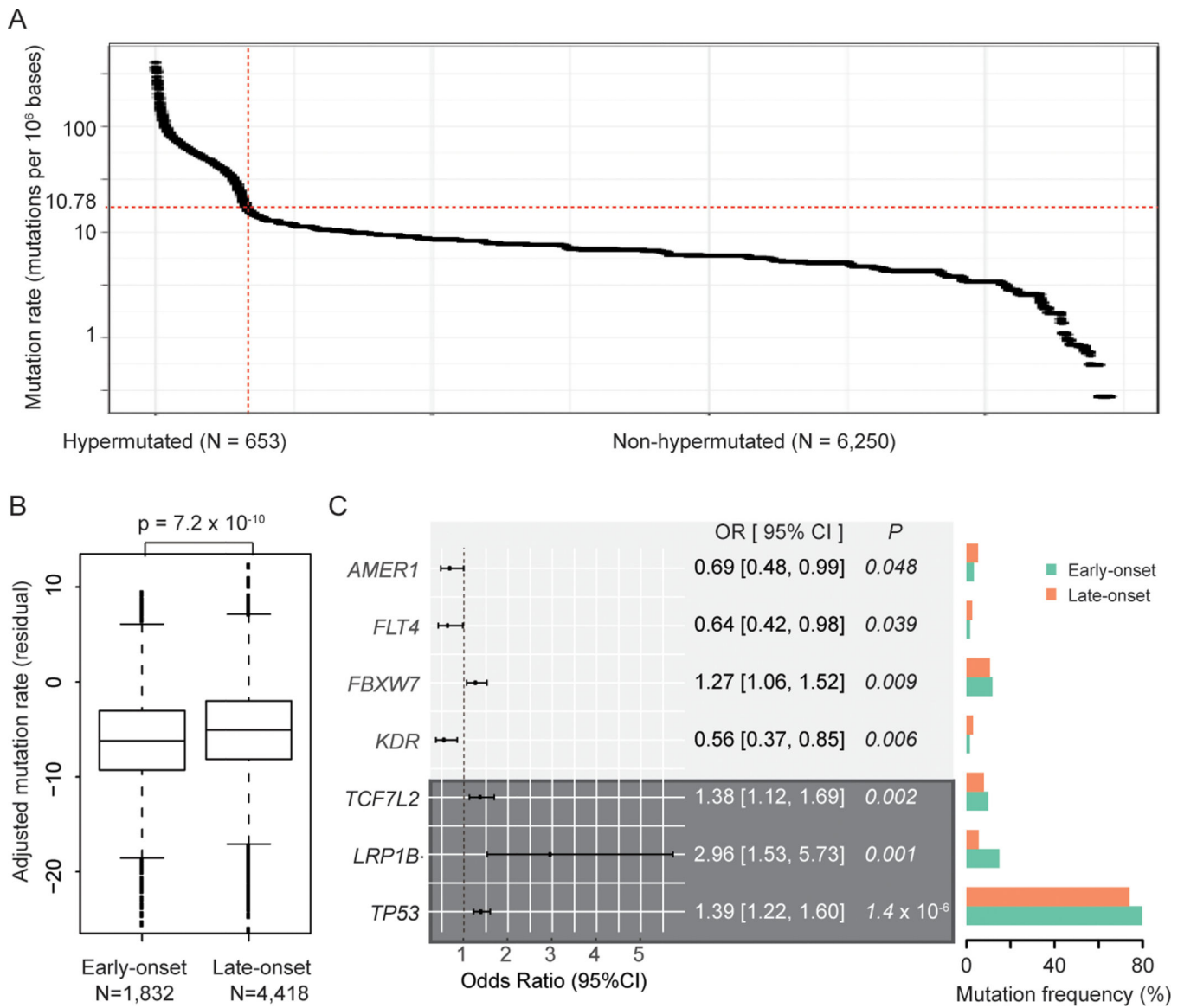


Figure 1. Genomic landscape of early-onset vs late-onset non-hypermutated colorectal cancer: AACR Project GENIE.

(A) Mutation rates among 6,903 tumor samples from CRC patients. Non-hypermutated tumors were defined using a cutoff (red line) of 17.78+ mutations/Mb.

(B) Boxplot of adjusted mutation rates between early-onset and late-onset cases with non-hypermutated CRC. The residual of adjusted mutation rates and *P*-value were derived from models adjusted for race and ethnicity, sex, tumor site and histology, sequencing assay, and sample type.

(C) Forest plot and mutation frequencies of genes differentially expressed between early-onset and late-onset non-hypermutated CRCs in adjusted models that reached statistical significance ($P < 0.05$). Genes with $FDR < 0.05$ are shaded in dark grey.

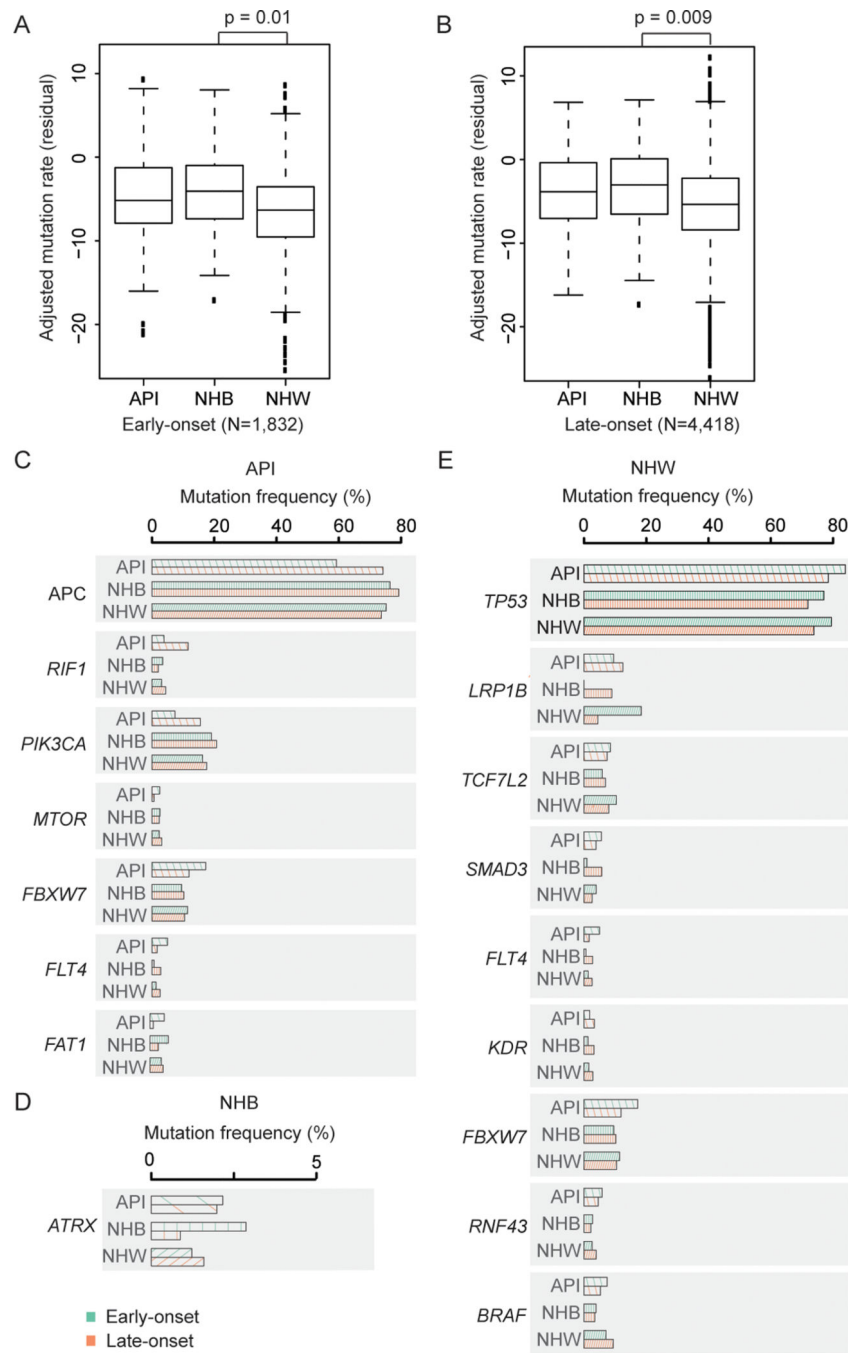


Figure 2. Racial/ethnic patterns of non-silent somatic cancer gene mutations among patients with early-onset non-hypermutated CRC.

API, Asian or Pacific Islander; NHB, non-Hispanic black; NHW, non-Hispanic white. (A-B) Boxplots of adjusted mutation rate residuals (tumor mutational burden, TMB) across racial/ethnic groups for (A) early-onset and (B) late-onset non-hypermutated CRC. The residual of adjusted mutation rates and *P*-values were derived from models adjusted for sex, tumor site and histology, sequencing assay, and sample type.

(C-E) Mutation frequencies between genes differentially expressed between early-onset vs late-onset non-hypermuted CRC cases that reached statistical significance for (C) API, (D) NHB, and (E) NHW patients.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

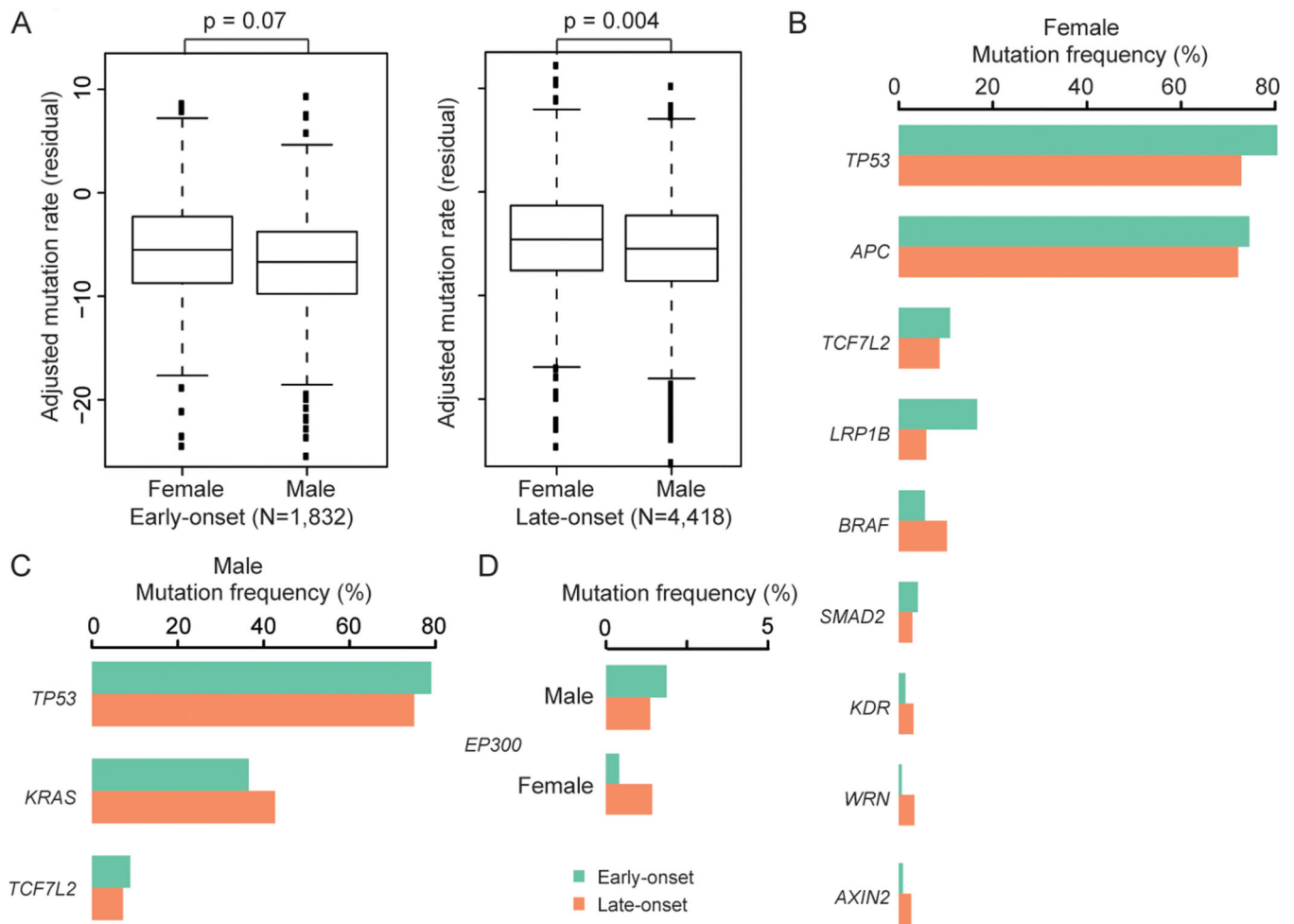


Figure 3. Tumor genomic profiles by sex among patients with early-onset non-hypermutated CRC.

(A) Boxplot of adjusted mutation rate residuals (tumor mutational burden, TMB) by sex for early-onset and late-onset non-hypermutated CRC. The residual of adjusted mutation rates and P -values were derived from models adjusted for race and ethnicity, tumor site and histology, sequencing assay, and sample type.

(B-C) Mutation frequencies between genes differentially expressed between early-onset vs late-onset non-hypermutated CRC cases that reached statistical significance ($P < 0.05$) for (B) females and (C) males.

(D) Inverse mutation frequencies for *EP300* in non-hypermutated CRCs among young patients by sex.

Table 1.

and heterogeneity of select non-silent somatic gene mutations by race and ethnicity among patients with early-onset colorectal cancer.

95% CI*	Non-Hispanic Black (NHB)										Asian or Pacific Islander (API)										Mutation frequency		Heterogeneity	
	FDR	Baseline mutation probability		Baseline mutation probability by age at cancer sequencing		95% CI*	OR*	P	FDR	Baseline mutation probability		Baseline mutation probability by age at cancer sequencing		95% CI*	OR*	P	FDR	P	Cochran's Q test	P-Het†				
		Early-Onset CRC	Late-Onset CRC	Early-Onset CRC	Late-Onset CRC					Early-Onset CRC	Late-Onset CRC	Early-Onset CRC	Late-Onset CRC											
1.20 - 1.61	0.001	0.74	0.77	0.72	1.31	0.80	2.14	0.29	0.995	0.80	0.84	0.78	1.66	0.97	0.06	0.60	0.25	0.49	0.78					
2.21 - 10.23	0.003	0.07	0.00	0.09	-	0.11	-	-	-	0.11	0.10	0.13	-	-	-	-	2.45E-10	-	-					
1.17 - 1.80	0.008	0.07	0.06	0.07	0.80	0.08	1.97	0.62	0.999	0.08	0.08	0.07	1.36	0.63	0.43	0.96	0.14	1.60	0.45					
1.14 - 2.71	0.18	0.04	0.01	0.06	0.25	0.04	2.05	0.20	0.995	0.04	0.06	0.04	1.65	0.58	0.35	0.93	0.07	3.16	0.21					
0.32 - 0.87	0.18	0.02	0.01	0.03	0.19	0.03	2.06	0.17	0.995	0.03	0.05	0.02	3.45	1.05	0.04	0.51	0.0006	9.19	0.01					
0.37 - 0.92	0.23	0.03	0.01	0.03	0.30	0.03	1.84	0.19	0.995	0.03	0.02	0.03	0.73	0.16	0.99	0.91	0.60	0.60	0.74					
1.04 - 1.54	0.23	0.10	0.10	0.10	0.90	0.14	1.80	0.76	0.999	0.14	0.17	0.12	1.86	1.03	0.04	0.51	0.04	2.53	0.28					
0.45 - 0.98	0.35	0.02	0.03	0.02	1.60	0.05	6.36	0.50	0.995	0.05	0.06	0.05	1.35	0.54	0.52	0.99	0.046	3.09	0.21					
0.63 - 0.99	0.35	0.04	0.04	0.04	1.21	0.06	3.53	0.72	0.999	0.06	0.07	0.05	1.15	0.49	0.75	0.99	0.27	1.23	0.54					
1.00 - 1.33	0.39	0.78	0.76	0.79	0.94	0.69	1.58	0.82	0.999	0.69	0.59	0.74	0.53	0.34	0.006	0.43	0.00001	10.58	0.005					
0.39 - 1.53	0.85	0.03	0.03	0.02	1.88	0.09	37.61	0.68	0.999	0.09	0.04	0.12	0.01	0.00	0.03	0.51	0.83	4.67	0.10					
0.86 - 1.20	0.93	0.20	0.19	0.21	1.00	0.13	1.70	0.99	0.999	0.13	0.07	0.16	0.47	0.23	0.03	0.51	0.003	4.49	0.11					
0.50 - 1.08	0.58	0.02	0.03	0.02	0.99	0.01	3.65	0.98	0.999	0.01	0.02	0.01	8.65	1.15	0.04	0.51	0.98	5.65	0.06					
0.66 - 1.32	0.92	0.04	0.06	0.03	2.35	0.02	7.28	0.14	0.995	0.02	0.05	0.01	4.37	1.02	0.047	0.51	0.32	6.07	0.048					
0.58 - 1.24	0.76	0.03	0.06	0.02	3.29	0.04	9.95	0.035	0.995	0.04	0.04	0.04	1.16	0.40	0.78	0.99	0.03	5.29	0.07					

CI, confidence interval; FDR, false discovery rate.

enes from models adjusted for patient sex, histology and site, sequencing assay, sample type and tumor mutational burden.

genity across the racial/ethnic groups. Only genes with significant associations for non-silent somatic mutations between early-onset and late-onset group were tested.