Integrative multi-omics profiling of colorectal cancer from a Hispanic/Latino cohort of patients

Waldrup B[1], Carranza F[1], Jin Y[1], Amzaleg Y[1], Postel M[2], Craig DW[1,3], Carpten JD[1,3], Salhia B[2,4], Ricker CN[4,5], Culver JO[4,5], Chavez CE[4], Stern MC[4,6], Baezconde-Garbanati L[4,6], Lenz HJ[4,5], Velazquez-Villarreal EI[1,3]

1. City of Hope, Beckman Research Institute, Department of Integrative Translational Sciences, Duarte, CA
2. University of Southern California, Keck School of Medicine of USC, Department of Translational Genomics, Los Angeles, CA
3. City of Hope Comprehensive Cancer Center, Duarte, CA
4. University of Southern California, USC Norris Comprehensive Cancer Center, Los Angeles, CA
5. University of Southern California, Keck School of Medicine of USC, Division of Medical Oncology, Los Angeles, CA
6 University of Southern California, Keck School of Medicine of USC, Department of Population and Public Health Sciences, Los Angeles, CA

ABSTRACT

Colorectal cancer contributes to cancer-related deaths and health disparities in the Hispanic and Latino community. To probe both the biological and genetic bases of the disparities, we characterized features of colorectal cancer in terms of somatic alterations and genetic similarity. Specifically, we conducted a comprehensive genome-scale analysis of 67 Hispanic and Latino samples. We performed DNA exome sequencing for somatic mutations, somatic copy number alterations, and genetic similarity. We also performed RNA sequencing for differential gene expression, cellular pathways, and gene fusions. We analyzed all samples for 22 important CRC gene mutations, 8 gene amplifications, and 25 CRC gene fusions. Then, we compared our data from the Hispanic and Latino samples to publicly available, Non-Hispanic White (NHW) cohorts. According to our analyses, twenty-four percent of colorectal carcinomas were hypermutated when patients were of Peruvians-from-Lima-like (1KG-PEL-like) genetic similarity population from the 1000 genome project. Moreover, most of these cases occurred in patients who were less than fifty years old age at diagnosis. Excluding hypermutated tumors, approximately 55% of colon cancers and 58% of rectum cancers exhibited two similar features: 1) the patterns of genomic alterations; 2) percentage of 1KG-PEL-like. We analyzed all samples -- which had a median 1KG-PEL-like proportion of 55% -- for 22 important CRC gene mutations, 8 gene amplifications, and 25 CRC gene fusions. One notable example of a frequently observed gene mutation was SMAD4. Samples with SMAD4 alterations, which are known to support tumor growth and progression, had the highest 1KG-PEL-like proportion (63%). According to our results from risk association analyses and differential gene expression, SMAD4 alterations were significant when we compared Hispanic and Latino samples to NHW cohorts. Of the 8 drug-targetable amplifications, PIK3CA and PI3K exhibited an average 1KG-PEL-like of over 55%. Of the 25 relevant CRC gene fusions, targetable genes included ALK, FGFR1, RAF1, and PTPRK; PTPRK was observed in a sample with the highest 1KG-PEL-like proportion (95%). Using integrative analysis, we also detected recurrent alterations in the WNT, TGFB, TP53, IGF2/PI3K, and RTK/RAS pathways. Importantly, these

alterations mostly occurred in young patients with high 1KG-PEL-like. These findings highlight the potential for tailoring precision medicine therapeutics to an underrepresented population. Our study advances the molecular profiling of CRC in Hispanics and Latinos. In toto, genetic similarity appears to be an important component in understanding colorectal carcinogenesis and has the potential to advance cancer health disparities research.

Significance: Our study provides molecular profiles of colorectal cancer in an underserved community, probes the impact of genetic similarity, and sheds light on cancer health disparities.

ABSTRACT (short version)
This study provides a comprehensive, genome-scale analysis of colorectal cancers (CRCs) taken from 67 Hispanic and Latino patients. We performed DNA exome sequencing for somatic mutations, somatic copy number alterations, and genetic similarity. We also performed RNA sequencing for differential gene expression, cellular pathways, and gene fusions. Our data from the Hispanic and Latino samples was compared to publicly available, Non-Hispanic White cohorts. In general, colorectal carcinomas from our cohort presented a high proportion (55%) of Peruvians-from-Lima-like (1KG-PEL-like) population from the 1000 genome project. Moreover, hypermutated colorectal carcinomas (24% of cases) predominantly occurred in individuals with 1KG-PEL-like. Most hypermutated cases occurred in young (<50) patients. We assessed our cohort for 22 important CRC gene mutations, 8 gene amplifications, and 25 CRC gene fusions. We cross-referenced results with percentage of 1KG-PEL-like. For example, frequent mutations were observed in a tumor suppressor called SMAD4; samples with SMAD4 alterations had 63% 1KG-PEL-like. Of the 8 drug-targetable amplifications, PIK3CA and PI3K exhibited an average 1KG-PEL-like of over 55%. Of the 25 CRC gene fusions, targetable genes included ALK, FGFR1, RAF1, and PTPRK. PTPRK was observed in a sample with 95% 1KG-PEL-like proportion. Using integrative analysis, we also detected recurrent alterations in the WNT, TGFB, TP53, IGF2/PI3K, and RTK/RAS pathways. Our study thus advances the molecular profiling of CRC in Hispanics and Latinos; suggests precision medicine therapeutics can be tailored to an underrepresented community; and demonstrates genetic similarity can be an important component in understanding colorectal carcinogenesis.

INTRODUCTION

Widely adopted screening programs, as well as greatly improved therapies, have decreased the overall burden of colorectal cancer (CRC).**[1-4][5]** Nevertheless, such progress has not been uniformly observed across all demographic groups.**[6-8]** Both the largest and fastest-growing minority group in the U.S. – the Hispanic and Latino community – continue to experience disparities in CRC incidence and outcomes.**[9]** The underlying reasons for these disparities are multifaceted: they involve complex interplay between socioeconomic factors, lifestyle, and genetics.**[10]** In emerging studies, these disparities appear to be linked to both genetic similarity and the specific molecular characteristics of tumors.**[11,12]** Although some studies have provided comprehensive, molecular characterizations of CRC, most of this work underrepresents the Hispanic and Latino population. They predominantly use data from Non-Hispanic White individuals. **[13,14]** They also typically categorize CRC tumors based on their hypermutated and

non-hypermutated status. For colorectal malignancies in the Hispanic and Latino community, little to no data is available about somatic mutations and genetic similarity. Altogether, significant gaps remain in our understanding of the molecular landscape of CRC in Hispanic and Latino individuals. Genetic similarity populations from the 1000 genome project of Hispanic and Latino populations plays a crucial role in understanding colorectal cancer tumorigenesis. **[15,16,17]** A study has shown that genetic similarity influences the molecular characteristics of CRC; genetic similarity affects both the mutation spectrum and the response to treatment.**[17]** For instance, certain genetic variants prevalent in admixed populations may contribute to higher rates of BRAF mutations in Hispanic/Latino populations. **[17]** Understanding these genetic differences is vital for addressing health disparities in CRC outcomes. Here, we aim to provide a comprehensive, genome-scale analysis of colorectal cancers (CRCs) taken from 67 Hispanic and Latino patients. We performed DNA exome sequencing for somatic mutations, somatic copy number alterations, and genetic similarity. We also performed RNA sequencing for differential gene expression, cellular pathways, and gene fusions. Our data from the Hispanic and Latino samples was compared to publicly available, Non-Hispanic White cohorts. We looked for several genes and pathways critical for the initiation and progression of CRC: six examples are WNT, RAS-MAPK, PI3K, TGF-β, P53, and DNA mismatch-repair **[15,16]**. Moreover, we assessed our data for identified recurrently mutated genes and a recurrent chromosomal translocation. The goal is to provide a deeper understanding of CRC disparities at the molecular level and identify potential targeted therapies for precision medicine.

RESULTS

Description of the study population

The study population for the Colorectal Cancer (CRC) Moonshot project consisted of 67 Hispanic and Latino patients (**Table 1**). Almost half of the patients (44.8%) were diagnosed under 50 years old. The gender distribution was similar: 53.7% male vs. 46.3% female. Most tumors (80.6%) were microsatellite stable (MSS) as opposed to microsatellite instable (19.4%, MSI). In terms of tumor mutational burden, fewer tumors (23.9%) were classified as hypermutated. The majority of tumors (62.7%) were located in the colon. Pathological staging revealed that 28.4% were in the early stages (Stage I or II), 43.3% were in the late stages (Stage III or IV), and 28.4% had no TNM applicable or an unknown stage.

Whole exome sequence (WES) analysis

To define the genetic similarity, mutational spectrum and chromosomal and sub-chromosomal changes, we performed whole exome sequencing on 67 tumors and normal pairs. Tumor and normal pairs were analyzed using the same platform. WES data summary quality metrics and statistics are provided in **Supplementary Table 1**.

Genetic similarity and tumor characteristics

Among our cohort, 100% self-reported as Hispanic and Latino. We deduced genetic similarity proportions and admixture for each patient using germline genetic similarity informative markers from WES data (**Fig 1a**). The mean of each genetic similarity population (1000 Genomes Project

Peruvian-in-Lima-like-1KG-PEL-like, 1000 Genomes Project European-like-1KG-EUR-like, 1000 Genomes Project African-like-1KG-AFR-like, 1000 Genomes Project South-Asian-like-1KG-SAS-like, 1000 Genomes Project East-Asian-like --1KG-EAS-like) was calculated according to participant status (**Table 1**). CRC cases had 58% 1KG-PEL-like, 36% 1KG-EUR-like, 4% 1KG-AFR-like, 1% 1KG-SAS-like, and 1% 1KG-EAS-like.

At the genetic similarity level within our cohort, all participants self-identified as Hispanic and Latino. Among hypermutated tumors, 100% showed a 1KG-PEL-like similarity of over 55% (**Fig. 1b**). In non-hypermutated tumors, 88% also exhibited a higher proportion of 1KG-PEL-like similarity, while 12% showed a 1KG-EUR-like similarity (**Fig. 1b**). These findings indicate a predominance of 1KG-PEL-like similarity in this population, as expected.

We measured the contribution of genetic similarity in the Hispanic and Latino population. Specifically, tumors from both 1KG-PEL-like and 1KG-EUR-like genetic similarity were characterized clinically and at the molecular level (**Supplementary Table 2**). Clinical characteristics included tumor location and tumor stage. Neither were significantly associated with either 1KG-PEL-like or 1KG-EUR-like. We also assessed the following 3 molecular tumor characteristics: microsatellite status, tumor mutational burden, and the mutation status of previously reported CRC genes in our study population. SMAD4-mutated tumors were found more frequently (~7 times) on patients with 1KG-PEL-like of >55%. However, the exact magnitude of risk increase is uncertain due to the wide confidence interval observed in the data (**Supplementary Table 3**).

Gene mutations

To define the mutational spectrum, we performed WES on 67 tumors and normal tissue. The somatic mutation rates varied considerably among the samples. Two samples had mutations rates of < 1 per 10^6 bases; two other samples had mutations rates of > 100 per 10^6. We separated cases (76%) with a mutation rate of < 9.23 per 10^6 (median number of total mutations, 91) and those with mutation rates of > 22.31 per 10^6 (median number of total mutations, 1852). The latter we designated as hypermutated **(Fig. 2a, Table 1)**. Compared to a previous study of CRC (N=224)**[13]**, we observed similar percentages of hypermutated and MSI tumors. Our data showed that 24% of the tumors (vs. 13% in **[13]**) were hypermutated. 81% of our hypermutated tumors (vs. 77% in **[13]**) exhibited MSI.

To assess the basis for the considerably different mutation rates —hypermutated (>22.31 per 10^6) and non-hypermutated (<9.23 per 10^6)—we evaluated MSI and mutations in the DNA mismatch-repair pathway genes MLH1, MLH3, MSH2, MSH3, MSH6 and PMS2. **[18, 19]** The following proportions of these 6 genes were observed in our cohort's 13 hypermutated MSI tumors: 38% MLH1, 31% MLH3, 23% MSH2, 62% MSH3, 46% MSH6, and 8% PMS2. These results agree with results reported by others in reference **[13].** Our study also identified similar POLE aberrations in the two samples with highest mutation rates and MSS. **(Fig. 2a)** All hypermutated samples exhibited high 1KG-PEL-like similarity (63% average). The majority were located in the colon (88%); 69% were early onset cases **(Fig. 2a)**.

In both the hypermutated and non-hypermutated cancers (our entire cohort), we identified 27 recurrent somatic mutations in CRC-related genes **(Fig. 2b)**. We observed 27 in hypermutated cancers and 21 in non-hypermutated cancers (**Fig. 2b**, for a complete list see **Supplementary Table 4**). Among the non-hypermutated tumors, the nine most frequently mutated genes were APC, TP53, KRAS, SYNE1, PIK3CA, TCFL2, OBSCN, AHNAK2, and FAT4. As expected, 34% of our samples, which had a higher 1KG-PEL-like (56% average), exhibited mutations in either KRAS (G12D, G12V, G12A, G12C, G13D,) or BRAF (V600E) (**Supplementary Table 5**); these mutations influence diverse mechanisms related to CRC treatment.**[20]** SACS, FAT3, SMAD4, FBXW7, BRAF and SOX9 were also mutated frequently. Mutations in SOX9, which is crucial for cell differentiation in the intestinal stem cell environment **[21,22]**, led to missense mutations, loss of function (LOF) mutations, multi-hit mutations, and copy number amplifications. The tumor suppressor genes we identified included APC, TP53, BRCA2, PTEN, FBXW7, NF1, and SMAD4. Among these, APC and TP53 showed the highest number of samples with mutations. NF1, PTEN, and SMAD4 were associated with the lowest average age at diagnosis, while PTEN and BRCA2 had the highest average of 1KG-PEL-like proportion (**Supplementary Table 6**). All these genes have been reported in CRC. **[13, 23]**

In the hypermutated tumors, we frequently observed mutations in TCF7L2, APC, SYNE1, OBSCB, SACS, AHNAK2, FAT3, FAT4, and BRCA2 (**Fig. 2b**). One gene was frequently mutated in non-hypermutated tumors but not in hypermutated tumors: TP53 (66 versus 31%, P<0.05). Twelve genes were more frequently mutated in the hypermutated tumors than in the non-hypermutated tumors: (1) TCF7L2 (94 versus 19% P=7.624e-08), (2) OBSCN (75 versus 13% P=), (3) SACS (75 versus 11% P=2.52e-06), (4) AHNAK2 (75 versus 11% P= 2.52e-06), (5) FAT3 (75 versus 9% P=9.036e-07), (6) FAT4 (63 versus 11% P= 9.104e-05), (7) BRCA2 (63 versus 4% P= 1.278e-06), (8) MSH3 (56 versus 2% P= 1.807e-06), (9) SOX9 (44 versus 6% P= 0.0008421), (10) PTEN (50 versus 4% P= 5.396e-05), (11) POLE (43 versus 2% P= 7.405e-05), and (12) NF1 (38 versus 2% P= 0.000404). Other genes, including MLH1, MSH2, MSH6, POLD1, RB1 and TET2, were mutated recurrently in the hypermutated cancers but not in the non-hypermutated samples. These findings indicate that hypermutated and non-hypermutated tumors progress through different sequences of genetic events. As expected, hypermutated tumors with MSI showed additional differences in the genomic and clinical profiles. All these tumors presented >55% 1KG-PEL-like proportion. Most of them were detected in colon from young patients who were less than 50 years old at diagnosis. (**Fig. 2b**).

Mutation rate

Colon and rectal tumors are managed differently in the clinic.**[24]** At the molecular level, we compared 42 colonic and 24 rectal tumors. We performed an initial, integrative analysis of MSI status, somatic copy-number alterations (SCNAs), genetic similarity status, and gene expression profiles. Our approach allowed us to probe potential biological differences between tumors in these two locations. As expected, in non-hypermutated tumors, the overall patterns of copy number changes, 1KG-PEL-like , and mRNA expression were indistinguishable between colon and rectal carcinomas (**Fig. 2a, Supplementary Fig. 1, Supplementary Fig. 2, Supplementary Table 7**). Based on these findings, we combined the two types (colon and rectal) for all subsequent analyses regardless of whether they were hypermutated or non-hypermutated.

Whole exome sequence comparison with public databases

We compared the frequencies of commonly mutated cancer genes from our CRC cohort to eight publicly available CRC genomic datasets: these included 148 cases from MSK-NatCommun, 1513 cases from MSK-JNCI, 619 cases from DFCI-CellReports, 224 cases from TCGA-Nature, 534 TCGA-PanCancerAtlas, 22 cases from MSK-CancerDiscovery, 471 cases from MSK-Gastroenterology, and 47 cases from MSK-JCO-PrecisOncol (**Table 2**).In nine datasets, including our study (**Supplementary Fig. 3**), the most commonly CRC mutated genes were APC, TP53, and KRAS. In two databases, these three genes consistently ranked among the top four most commonly mutated genes. Five frequently mutated genes in our Hispanic and Latino cohort were relatively distinct from some of the publicly available datasets: SYNE1, TCF7L2, OBSCN, PIK3CA, and SACS. For example, the percentage of SYNE1 mutations in our cohort was 38.8% vs. 27% in DFCI-CellReports , 21% in TCGA-Nature, 27% in TCGA-PanCancerAtlas. Mutations in OBSCN 28.36% were more frequent in the Hispanic and Latino cohort than three databases with ranges from 6.30% to 18.70%. While the percentage of PIK3CA mutations in our cohort (28.36%) was similar to six databases, it differed drastically from one database with a considerably higher percentage (40.90%) and one with a considerably lower percentage (6.10%). The percentage of tumors with mutated SACS in the Hispanic and Latino cohort (26.87%) was higher than three databases (9.80%, 10.80%, and 12.50%).

Of 19 genes we studied in our cohort, sixteen of them exhibited similar percentages of mutations to at least one of the eight public databases. Three genes had slightly higher frequencies than all the referenced databases (**Table 2**). Two genes exhibited high, statistically significant differences with respect to the largest dataset (MSK-JNCI): TCF7L2 (p-values of 6.725E-07) and BRCA2 (p-value of 8.45E-05). Both of these genes had higher mutational frequencies in our cohort (**Table 2**). Other genes with notable variant differences with respect to the largest dataset include TP53, MSH3, PTEN, MLH1, MSH6, RB1 and MSH2 (**Table 2**). Ethnicity information was reported only in the TCGA-PanCancer Atlas: an extremely small percentage of samples (0.8%) of the samples identified as Hispanic and Latino.

Furthermore, we compared our CRC cohort to two previous studies—Innocenti et al. (2023, *JCO*) and Nunes et al. (2024, *Nature*)—as detailed in Supplementary Table 8. Notably, one of these studies highlighted significant differences in the frequency of APC and KRAS mutations when compared to our cohort, particularly among non-Hispanic White (NHW) individuals. In our Hispanic and Latino cohort, seventeen genes were frequently mutated and showed relative distinction from one or both studies. These genes include: TCF7L2, PIK3CA, FAT3, BRCA2, SOX9, PTEN, POLE, NF1, MLH1, MSH6, SMAD4, BRAF, RB1, MSH2, TET2, POLD1, and MUTYH. This distinct mutational landscape underscores the potential genetic differences in CRC across diverse populations.

We additionally assessed somatic mutations in CRC cases from the AACR Project GENIE dataset (**Table 3**). This dataset includes self-identified race and ethnicity information. We combined cases from our cohort, TCGA-PanCancer and AACR Genie, while stratifying by self-reported race. This combined dataset included a total of 140 (9%) patients who self-identified as

Hispanic and Latino. The remaining 1487 (91%) cases self-identified as Non-Hispanic White (**Table 3**). Based upon this large cohort, which was race stratified by self-identification, we identified twelve statistically significant differences in the frequency of commonly mutated genes: SYNE1, TCF7L2, OBSCN, PIK3CA, SACS, AHNAK2, FAT3, FAT4, BRCA2, MSH3, POLE and RB1 (**Table 3**).

Chromosomal and sub-chromosomal changes

In total, 67 tumors were profiled by Whole Exome Sequencing (WES) for Somatic Copy-Number Alterations (SCNAs, **Fig. 3**). As previously reported in CRC studies, the hypermutated tumors had far fewer SCNAs (**Supplementary Fig. 4**) than non-hypermutated tumors. However, hypermutated MSI samples had more SCNAs than MSS samples (**Supplementary Fig. 5**). In addition, samples with ≤55% 1KG-PEL-like proportion had slightly more amplifications than samples with >55 1KG-PEL-like proportion (**Supplementary Fig. 6**). No significant differences were detected between non-hypermutated colon and rectum (**Supplementary Fig. 1**).

We used the Genomic Identification of Significant Targets in Cancer (GISTIC) algorithm and manually curated data to identify probable gene targets of focal alterations.**[25]** Most tumors had several, previously well-defined arm-level changes: gains of 1q (93%), 7p (96%) and q (92.5%), 8p (87%) and q (97%), 12q (87%), 13q (96%), 19q (93%), and 20p (97%) and q (100%) (**Supplementary Table 9**). Significantly amplified chromosomes arms were 1q (including OBSCN) in 2% of the samples, 13q (including SACS, BRCA2 and RB1) in 10%, and 14q (including AHNAK2) in 3% (**Supplementary Fig. 8**). CCND1 and OBSCN had also deletions as described below. Other significantly deleted chromosome arms were 2p (80%), 13p (99%), 14p (100%), 15p (100%), 20q (80%), 22p (96%) (**Supplementary Table 9**). Compared to previously reported**[13]**, CRC-related deletion of chromosome arms, our samples exhibited smaller amounts of deletions in 1p (28%), 4q (55%), 5q (30%), 8p (48%), 14q (40%), 15q (48%), 20p (78%) and 22q (49%) (**Supplementary Table 9**). Other significantly deleted chromosomes arms were 1p and q (including MUTYH in 2% and OBSCN in 7% of the samples), 4q (including FAT4 in 2% and FBXW7 in 2%), 5q (including APC in 5% and MSH3 in 2%), 10q (including PTEN in 6%), 11q (including FAT3 in 3%), 12q (including POLE in 3%), 14q (including AHNAK2 in 2%), 18q (including SMAD4 in 12%), and 19q (including POLD1 in 2%) (**Supplementary Fig. 7**).

Our data from the 67 Hispanic and Latino patients had 66 recurrent deletion peaks (**Supplementary Table 10a**). The following four categories of genes were affected by focal deletions in our samples: (1) RBFOX1 and WWOX;**[13]** (2) the tumor-suppressor genes SMAD4**[26]** and APC;**[27]** (3) the mismatch repair genes MSH2 and MSH6**;[28]** and (4) a ubiquitin ligase called FBXW7, which targets oncoproteins for degradation (alterations can lead to the accumulation of oncogenic proteins).**[29]** Our data set also had a significant focal deletion of 1q42.13, which spanned two genes. This included OBSCN, which was also frequently mutated in our data set. Other significant focal deletions were 4q28.3, 12q24.33, 4q28.3, and 19q13.41: they involved the genes FAT4, POLE, TET2A and POLD1. Four significant focal deletions 6q25.2, 2p23.1, 2p11.2, and 18q21.2 affect genes detected in our gene fusion analyses: PTPRK, ALK, EML4 and RSPO3. While these appear to be non-significant, we observed 25 regions of significant focal amplification of 6p22.2; this included two genes – HIST1H4H **[28]** and BTN3A2 **[30,31]** – which play crucial roles

in both histone modification patterns and regulation of immune responses (**Supplementary Table 10b**). Some of these overlapped with extensive gains on chromosome arms and featured notable peaks at 8q24.21 and 7p11.2. Importantly, these peaks encompass both the key oncogene MYC **[32]** and the well-known, druggable target gene EGFR.**[33]**

Transcriptomics Analysis

We performed RNA sequencing on the same 67 tumors to define gene expressions profiles and fusions and pathways.

Transcriptomic comparisons with public databases: Hispanics and Latinos compared to Non-Hispanic Whites

Compared to colorectal tumors of Non-Hispanic White populations, tumors from our Hispanic and Latino patients had unique gene expression patterns and cellular pathways. Specifically, we conducted a differential gene expression and pathway analysis using RNA sequencing (**Fig. 4, Supplementary Fig. 8**). We compared gene expression profiling in 67 tumors from the Hispanic and Latino patients to 67 tumors from Non-Hispanic White patients. Data from Non-Hispanic White patients were taken from the TCGA-COAD and TCGA-READ projects. We matched tumors from the different populations using parameters shown in Table 1: tumor location, MSI status, age at onset, and TNM classification (**Supplementary Table 11**). A comparison of case matched ratios revealed at least 10 statistically significant CRC related genes at p-value of <0.0001. These genes were as follows: PIK3CA, MUTYH, PMS2P3, POLE, PTEN, SYNE1, POLD1, SMAD4, HFM1, and FAT3. Pathway analysis identified 10 cellular pathways related to colorectal cancer. These pathways were related to dysregulation of GTPases, microRNA (miRNA), non-coding RNAs, post-transcriptional mechanisms, apoptotic signaling, mRNA degradation, ribosome biogenesis, protein-protein interactions, and mitochondrial membrane organization. Notably, activation of GTPase activity, such as RAS and RHO family members, is commonly associated with colorectal cancer.

Hypermutated and non-hypermutated tumors

Using the same approach (*vide supra*), we compared whether hypermutated samples differed from non-hypermutated samples (**Supplementary Fig. 9**). Differential gene expression analysis revealed 9 statistically significant CRC related genes at p-value of <0.05. These genes were MSH4, TET2, FAT4, TCF7L2, HFM1, SACS, OBSCN, SYNE1, and SMAD4 (**Supplementary Table 12**). Among the 20 top statistically significant pathways, we identified at least 6 cellular pathways implicated in CRC. These pathways were involved in dysregulations of protease activity, endopeptidase activity, calcium signaling, hydrolase activity, proteolysis activity, and catalytic activity. Notably these pathways are directly or indirectly relevant to CRC biology and may influence tumor growth, invasion, and metastasis.

Higher proportions of 1KG-PEL-like

As above (*vide supra*), we compared tumors with > 55% 1KG-PEL-like to those with ≤ 55% (**Fig. 4c, Supplementary Fig. 10**). The differential gene expression revealed two statistically significant CRC related genes at p-value of <0.0001. These genes were TET2 and HFM1 (**Supplementary Table 13**). Among the 20 top statistically significant pathways, we identified at least 6 cellular pathways

associated with CRC. These pathways were involved in dysregulated cell growth, cellular pathways, and immune responses. Notably, these pathways are either directly or indirectly relevant to CRC biology and may influence tumor growth, invasion, and metastasis.

Gene fusions

While gene fusions are atypical in CRC, the emergence of fusion-targeted therapeutics underscores the crucial need for identifying clinically actionable fusions. **[34-36]** To identify gene fusions, we performed gene fusion analysis using star fusion in all the tumors (**Fig. 5 and Supplementary Table 14**). With this high genome coverage, we detected in average 2 gene fusion events per sample. Among these events, we identified 21 samples with gene fusions. Fusions typically involved one of the following three partners: either Mismatch Repair System Component Pseudogene 9 (PMS2P9), Mismatch Repair System Component Pseudogene 11 (PMS2P11), or Mismatch Repair System Component Pseudogene 9 (PMS2P9). One of these were typically fused with either Coiled-Coil Domain Containing 146 (CCDC146) or AC105052.5.**[37, 38]** Most of these fusions were more prevalent in samples with MSS (90%) than in MSI (10%). Most also occurred in patients with 1KG-PEL-like > 55%.

TP53I11 is a gene under the regulatory control of TP53. In four samples, TP53I11 was fused with OR51S1; these samples also had MSS and ~~mostly~~ had 1KG-PEL-like > 55%. Additionally, our analysis revealed two clinically actionable gene fusions in separate samples of CRC: (1) Anaplastic Lymphoma Kinase (ALK) and (2) Fibroblast Growth Factor Receptor (FGFR). The sample with the ALK fusion was MSI; the sample with the FGFR fusion was MSS. Both samples had 1KG-PEL-like >55%. Importantly, CRCs with these fusion have been successfully treated with ALK and FGFR inhibitors.**[39-42]** Additionally, two patients in our cohort exhibited gene fusions relevant for CRC. (1) One colon tumor sample harbored the PTPRK-RSPO3 gene fusion. This young patient (<50) exhibited MSS and an 1KG-PEL-like >55%. The PTPRK-RSPO3 gene fusion plays a role in both activating the WNT signaling pathway,**[43, 44]** and promoting CRC. (2) One rectum tumor sample harbored the RAF1-TMEM40 gene fusion. This case was late-onset (≥50yrs), exhibited MSS, and was 1KG-PEL-like of >55%. RAF1 gene fusions, such as RAF1-TMEM40 (seen in our samples), are part of the broader RAF kinase gene family; this includes BRAF, RAF1, and CRAF. These gene fusions are known to activate the MAPK signaling pathway and have been found across various types of cancers. RAF1 fusions can contribute to oncogenic activity in CRC. **[45]**

Altered pathways

The 67 samples underwent integrated analysis for mutations, copy number, mRNA expression, and genetic similarity in 67 tumors. Our objective was to understand of how some well-defined pathways were deregulated. We grouped samples by hypermutation status and identified recurrent alterations in WNT, TGFB, TP53, IGF2/PI3K, RTK/RAS, and combined RTK/RAS/PI3K pathways (**Fig. 6**).

Ninety-six percent of tumors (64/67) had perturbations in the WNT signaling pathway; accordingly, a previous study (N=165) reported such perturbations in 93% of tumors. In our cohort, 48% of the tumors (31/64) had an 1KG-PEL-like >55%. Fifty-three percent (36/67) were

early onset. As expected, we found higher mutations (80%) than SCNA (10%) in APC. CTNNB1 exhibited three times as many mutations (12%) as SCNAs (4%). Both mutations and SCNAs were found in SOX9 and TCF7L2, as well as the DKK family members and AXIN2, FBXW7 and ARID1A **[46, 47]**. In the RNAseq analysis results we obtained, the typically considered tumor suppressor genes ARID1A and AXIN2 were downregulated (meaning their expression levels were decreased), while the oncogene CTNNB1 was upregulated (meaning its expression level was increased) **Transcriptomics comparison with public databases: Hispanics/Latinos and Non-Hispanic Whites**). Altogether, we observed expression differences between the Hispanic and Latino samples vs. the Non-Hispanic White samples (Fig 4). In the Hispanic and Latino cohort, various genetic alterations were observed, including mutations, deletions, amplifications and dysregulations in WNT-related genes such as TCFL2, SOX9, LRP5, FZD10, FBXW7, DKK4, DKK3, DKK2, DKK1, CTNNB1, AXIN2, and ARID1A, alongside APC mutations in tumors (**Fig. 6**). Thus, this tumor suppressor gene plays a crucial role in disrupting the WNT signaling pathway.

Genetic alterations in both the PI3K and RAS pathways are common in CRC. According to our results, 86% of tumors had perturbations in the combined RTK/RAS/PI3K pathway. From these, 47% had an 1KG-PEL-like of >55% and 55% were early onset. We found higher number of mutations and SCNAs (amplifications and deletions) in IRS2 than in IGF2. PIK3CA exhibited nearly four times more mutations than PIK3R1. PIK3CA also exhibited amplifications but no deletions. Conversely, PIK3R1 had deletions but no amplifications. Compared to tumors with 1KG-PEL-like <55%, slightly more tumors with 1KG-PEL-like >55% presented KRAS mutations and double the number of the BRAF mutations. Interestingly, none of the tumors presented NRAS mutations. PTEN deletions were present in tumors with different 1KG-PEL-like proportions. We also investigated mutations in the erythroblastic leukemia viral oncogene homolog (ERBB) family of receptors due to their significant translational implications. Specifically, ERBB2 and ERBB3 mutations were identified in 16% and 7% of the tumors, respectively, with the majority exhibiting 1KG-PEL-like proportions <55%. Oncogenes PIK3CA, NRAS, KRAS and BRAF were significantly upregulated. Compared to a previous study **[13]**, we identified slight differences in perturbations ~~alterations~~ involving RAS and PI3K pathways. Altogether, common CRC pathways exhibited genomic heterogeneity for our Hispanic and Latino cohort.

The TGF-b signaling pathway is known to be deregulated in CRC and other cancers **[48]**. In 70% of the 67 samples, we found genomic perturbations in TGFBR1, TGFBR2, ACVR2A, ACVR1B, SMAD2, SMAD3 and SMAD4. In this 70%, almost half of the samples had 1KG-PEL-like >55% and were early onset. Consistent with a previous study **[12]**, we found alterations in TP53 in 63% of the tumors. Additionally, we observed more than double the expected alterations in ATM, a kinase that phosphorylates and activates P53 after DNA damage; 19% of the tumors exhibited these alterations. Interestingly, the tumor mutation gene TP53 was significantly downregulated in the Hispanic and Latino population vs. the Non-Hispanic White population (Fig. 4).

DISCUSSION

This study provides a comprehensive integrative analysis of 67 colorectal tumor and germline matching samples from Hispanic and Latino individuals. It yielded several key insights into the

biology of colorectal cancer (CRC) and has highlighted potential therapeutic targets. For example, we observed interesting differences between the hypermutated (MSI) and non-hypermutated tumors (MSS). For example, the non-hypermutated samples were predominantly located in the left colon. Some of these samples exhibited high 1KG-PEL-like while others had mostly 1KG-EUR-like. In contrast, the hypermutated samples were mainly found in the right colon and exhibited predominantly 1KG-PEL-like. (**Fig. 1**). We do not know why most of the hypermutated samples came from the right colon and exhibited predominantly 1KG-PEL-like. Given that patients with MSI-related cancers, which are hypermutated, tend to have better survival rates, the mutation rate might serve as a more effective prognostic indicator. Additionally, the high prevalence of 1KG-PEL-like in these samples could also serve as another important prognostic factor.

In general, the percentages of genetic similarity populations were not associated with CRC tumor mutations. However, one exception was the SMAD4 gene; tumors with mutated SMAD4 had high (>55%) 1KG-PEL-like. Previous studies have demonstrated associations between genetic similarity differences and various cancers, including breast, melanoma, and prostate cancer.**[49-56]** Two prior studies examined the association of genetic similarity with CRC in Hispanic and Latino populations – in Puerto Ricans**[11]** and Colombians**[57]**. Their results suggest slight differences primarily in the 1KG-AFR-like component. In our study, SMAD4-mutated tumors were significantly more frequent—almost seven times higher—in patients with an 1KG-PEL-like proportion of >55%. However, the exact magnitude of the risk increase remains uncertain due to the wide confidence interval observed in the data. The high prevalence of SMAD4 mutations in our cohort of Hispanic and Latino individuals with CRC in Los Angeles is concerning due to its association with poor prognosis and treatment resistance. SMAD4 is a crucial tumor suppressor gene involved in the TGF-beta signaling pathway, and its loss of function is linked to more aggressive tumor behavior, lower survival rates, and resistance to standard therapies like EGFR inhibitors and anti-VEGF therapies **[86-88]**. These challenges necessitate personalized treatment approaches for SMAD4-mutated tumors. Moreover, our findings suggest that genetic disparities, such as the predominance of 1KG-PEL-like genetic similarity, may also contribute to CRC disparities in this population, alongside socioeconomic factors **[89]**. Thus, understanding the interplay between genetic and socioeconomic factors is essential for developing targeted interventions and improving outcomes for Hispanic and Latino CRC patients.

Our study compared the mutation frequencies of commonly mutated cancer genes in our CRC cohort with those from eight publicly available CRC genomic datasets: MSK-NatCommun, MSK-JNCI, DFCI-CellReports, TCGA-Nature, TCGA-PanCancerAtlas, MSK-CancerDiscovery, MSK-Gastroenterology, and MSK-JCO-PrecisOncol **(Table 2 and Fig. 4)**. Across all datasets, APC, TP53, and KRAS were the most frequently mutated genes. These three genes were also the most frequently mutated in our cohort. Interestingly, the mutation frequency of SYNE1 in our Hispanic and Latino cohort was slightly higher than in other databases. TCF7L2 and OBSCN showed comparable or slightly elevated mutation frequencies. Our data revealed that 16 out of 19 genes had similar mutation frequencies to at least one of the public databases, with TCF7L2 and BRCA2 exhibiting significantly higher mutation frequencies compared to the MSK-JNCI dataset. We used the AACR Project GENIE dataset, which is stratified by self-reported race, to compare the mutational frequencies of the Hispanic and Latino cohort to those of the Non-Hispanic White

population: we observed significant differences in mutational frequencies of TCF7L2, PIK3CA, SACS, AHNAK2, MSH3, and RB1. Although our comparisons with large-scale analyses revealed notable similarities, they also uncovered distinct molecular characteristics in the Hispanic and Latino population. Altogether, diverse populations should be more comprehensively included in genetic studies to better understand CRC disparities and identify potential targets for precision medicine.

Whole-exome sequencing and integrative analysis of the genomic data provided further insights into the pathways affected in CRC when comparing Hispanic and Latino and Non-Hispanic White populations. For example, 96% of the tumors in our cohort exhibited mutations and SCNAs in the WNT signaling pathway. Compared to gene expression in the Non-Hispanic White cohort, our cohort had unique, deregulated genes in the WNT signaling pathway: two examples are the tumor suppressor genes ARID1A and AXIN2. ARID1A has been shown to suppress MYC transcription and to develop a mechanism of resistance to cetuximab.**[58, 90]** MYC activation is known to result from activation of the WNT signaling pathway and inactivation of the TGF-β signaling pathway. Accordingly, we found alterations in TGF-β signaling genes in 70% of the tumors, which suggested a significant role of MYC in CRC among Hispanic and Latino patients that merits further study. Additionally, when we used differential gene expression analysis to compare our cohort with the Non-Hispanic White population, we observed that the oncogene CTNNB1 was upregulated.[**59,60**] CTNNB1 encodes the protein β-catenin, a crucial component of the WNT signaling pathway. Another affected component of WNT signaling was TCF7L2: tumors with 1KG-PEL-like >55% had more than double TCF7L2 mutations than tumors with 1KG-PEL-like <55. Altogether, our integrative analysis underscores the pivotal role of the WNT signaling pathway in the oncogenesis of CRC in this Hispanic and Latino patients.

According to our results, colon and rectum tumors have biological similarities which, to the best of our knowledge, were previously unknown. These similarities are important because colon and rectal tumors do not have the same standard-of-care. Irrespective of their anatomical origin, non-hypermutated tumors from our cohort had similar types of copy numbers and expression profiles. More than 90% had perturbations in one or more members of the WNT signaling pathway; all these samples exhibited predominantly 1KG-PEL-like. Additionally, from a therapeutic perspective, our data from two samples provide clinically actionable gene fusions involving ALK and FGFR. Targeted therapies, including ALK and FGFR inhibitors, have been successfully implemented in the treatment of CRC following the detection of these fusions.**[39, 40]** Another clinical actionable approach is the WNT-signaling inhibitors and small-molecule b-catenin inhibitors.**[61-63]** We find that several proteins in the RTK–RAS (drug targetable fusions) and PI3K pathways, including IGF2, ERBB2, and ERBB3 could be targets for inhibition.

CONCLUSION

The data presented in this study provide a valuable resource for understanding the molecular characteristics of CRC in the context of 1KG-PEL-like proportions and potential therapeutic targets in an underrepresented minority group. This is the first integrative report to examine the multi-

omics landscape and the associations of genetic 1KG-PEL-like with CRC in Hispanic and Latino populations, compared to multiple studies in Non-Hispanic White. Our findings suggest potential differences in the disease's biology, which could influence screening and treatment options for patients. Further studies with larger sample sizes are needed to elucidate the genomic heterogeneity and the role of 1KG-PEL-like in the development of CRC tumors in Hispanic and Latino populations.

MATERIALS AND METHODS

Study population

Participants in this study were recruited through the NIH NCI grant project number U2CCA252971, a Center for Optimizing Engagement of Hispanic Colorectal Cancer Patients in Cancer Genomic Characterization Studies in the University of Southern California, Los Angeles area, that started in 2021. The University of Southern California COPECC center (https://usccopecc.org) is a member of the Participant Engagement and Cancer Genome Sequencing (PE-CGS) Network (https://pe-cgs.org) part of the Cancer Moonshot Initiative^{SM} (https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative) that is intended to accelerate cancer research.

Data collection

All participants completed questionnaires as part of the USC COPECC center Patient Engagement Unit (https://usccopecc.org/patient-engagement-unit/) which the main goal is to develop a patient engagement framework uniquely tailored for Hispanic cancer patients.

Clinical data

Clinical data is obtained for all cases through USC COPECC center. Data on tumor location, tumor stage (TNM stage), MSI, primary site of diagnosis, colon side, pathological group stage at diagnosis, and medications used were collected.

DNA and RNA extraction and quality control

Tumor tissue and blood samples were collected from study participants according to standard operating procedures through the USC COPECC center Genome Characterization Unit (https://usccopecc.org/genome-characterization-unit/).

DNA Isolation

Genomic DNA and RNA was extracted from tissue using the Covaris truXTRAC© total NA Column kit (Woburn, MA) based on the manufacturer's recommendations. Briefly, tissue samples were placed into Covaris Adaptive Focused Acoustics (AFA) and suspended in Lysis buffer followed by Proteinase K digestion. Tissue mixtures were then equally split for separate DNA and RNA extraction. Tissue mixtures were then ultrasonically emulsed using the Covaris E220 System. DNA or RNA sample emulsions were then separately chemically de-crosslinked and pipetted onto appropriate spin columns. DNA and RNA were then collected using provided elution buffers. DNA and RNA were quality assessed and quantified using both the (Thermo Fisher Scientific, Inc.

Waltham, MA) and the Genomic DNA Screen Tape Assay utilizing the Agilent 4200 TapeStation System (Santa Clara, CA). DNA and RNA samples were stored at -80°C.

Whole exome and RNA sequencing

For Whole exome sequencing we utilized a custom expanded exome bait set (Agilent Technologies, Inc.). Briefly, components of the expanded exome included the following probe groups: original baits from SureSelect Human All Exon V6, (Agilent Technologies, Inc.) and custom baits for select genomic regions.[**64**] Genomic DNA in the amount of 50-200 nanograms (ng) from colorectal tumor tissue and paired uninvolved (normal) colorectal tissue from each case was sheared in 50 microliters (μl) of TE low EDTA buffer employing the Covaris E220 system (Covaris, Inc., Woburn, MA) to target fragment sizes of 150 – 200 bp. Fragmented DNA was then converted to an adapter-ligated whole genome library using the Kapa Hyper Prep Library Prep kit (Kapa Biosciences, Inc., Wilmington, MA) according to the manufacturer's protocol. SureSelect XT Adaptor Oligo Mix was utilized in the ligation step (Agilent Technologies, Inc.). Individual tumor adapter-ligated libraries were enriched into the exome capture reaction, and for germline each adapter-ligated library was pooled before proceeding to capture using Agilent's SureSelect Human All Exon V6 + custom probes capture library kit. Samples that had successful libraries created were then sequenced on Illumina MiSeq technology for quality control to assess the ability of the libraries to be sequenced. Subsequently, each library was pooled and sequenced on Illumina's NovaSeq 6000 (Illumina, San Diego, CA) using 300 cycle kit. Raw FASTQs were generated using the industry standard BCL2FASTQ v1.8.4. Mean target coverage was 105X for tumor samples and 55X for uninvolved control samples.

RNA was extracted from cells from colorectal cancer tissues samples and analyzed by RNA-seq. Sequencing libraries prepared with the TruSeq Stranded Total RNA kit (Illumina Inc), from 1 μg total RNA.

Whole exome sequencing analysis

All sequencing reads were converted to industry standard FASTQ files using the Bcl Conversion and Demultiplexing tool (Illumina, Inc). Sequencing reads were aligned to the GRCh38 reference genome using the MEM module of BWA v0.7.17[**65**] and SAMTOOLS v1.9 [117] to produce BAM files. After alignment, the base quality scores were recalibrated and joint indel realignment was performed on the BAM files using GATK v4.0.10.1.[**66**] Duplicate read pairs were marked using PICARD v2.18.22.[**67**] Final BAM files were then used to identify germline and somatic events. Germline SNP and INDELS were identified using GATK haplotype caller in the constitutional sample.

Gene mutations analysis

Somatic variant callers identified single nucleotide variants (SNVs) and small insertions and deletions (indels). Somatic mutation analysis was performed by identifying overlapping variant calls generated by Strelka[**68**] and Mutect.[**69**] Due to the diffuse nature of these tumor samples, we retained somatic mutations that had a minor allele frequency of 4%, and manually reviewed variants in the KRAS region using IGV.[**70**] We compared these mutations against germline results to confirm and verify the somatic mutations. After filtering and manually reviewing somatic variants, VCFs were then annotated with Ensembl's VEP[**71**] tool and converted to the MAF file

format for visualization purposes. MAFtools[**72**] was utilized in the generation of oncoplots, annotated with demographic information associated with each sample.

Somatic Copy-Number Alterations (SCNAs) analysis

For chromosomal and sub-chromosomal changes analysis we used GISTIC2.0 **[25]** best practices to identify genes targeted by somatic copy-number alterations (SCNAs) and SCNA profiles into underlying arm-level and focal alterations.

Genetic Similarity analysis

We used GATK HaplotypeCaller v4.0.10.1 to generate germline VCF files for genetic similarity analysis. Known population genotype data were downloaded from the 1000 Genomes Project phase 3 and grouped by super population for genetic similarity analysis.[**73**] VCF subsetting and merging were performed using VCFtools v0.1.17 and SnpSift v4.3t.[**74**,**75**] VCF genotype allele coding (0/1) was converted to numeric (012) using VCFtools and PLINK v1.90b6.7.[**76**] All genetic similarity analyses were performed on autosomal chromosomes. The population admixture was estimated using STRUCTURE v2.3.4.[**77**] We subset and merged all VCF files by a list of 1766 genetic similarity-informative markers. The result was converted to Structure format by PLINK. To run STRUCTURE, we set up population numbers k=5, NUMREPS=2000, and BURNIN=50000. Principal component analysis (PCA) was performed on the same dataset by R functions[**78**]. All similarity analysis results were visualized using R v3.6.0 ggplot2 v3.4.1[**79**] package. For initial analysis, genotypes for a total of 1,766 genetic similarity-informative markers were extracted and used for estimating genetic similarity based upon five genetic similarity super continental populations including 1000 Genomes Project African-like (1KG-AFR-like), 1000 Genomes Project Peruvian-in-Lima-like (1KG-PEL-like), 1000 Genomes Project European-like (1KG-EUR-like), 1000 Genomes Project East-Asian-like (1KG-EAS-like), and 1000 Genomes Project South-Asian-like (1KG-SAS-like) using STRUCTURE (**Fig. 1**).

RNA sequencing analysis

Alignment of raw sequence reads to the human transcriptome (hg38) was performed via Rsubread[**80**] and transcript abundance estimates were normalized and differentially expressed genes (DEG) identified using a standard edgeR pipeline. Functional annotation of gene sets: Pathway enrichment analysis and gene set enrichment analysis (GSEA) were performed using gene sets from the Molecular signatures database (MSigDB). In addition, R package gage (Generally Applicable Gene-set Enrichment) where used to identify enriched pathway and pathview to visualize these pathways. For transcript-aware analyses, the FASTQ files were aligned with salmon[**81**] and differentially enriched transcripts were identified using DRIMSeq[**82**] in a similar workflow to edgeR[**83**].

Gene fusions analysis

For gene fusion we used STAR-Fusion that is built on top of the STAR aligner according with its best practices.[**84**] We ran STAR version 2.5.3a, STAR index was created, and reads were aligned to the same human reference genome than used for the WES analysis.

Statistical analysis

In this study, we included 67 subjects recruited from USC COPECC, of which all passed the genomic sequencing quality controls (**Supplemental Table 1**). Clinical and demographic data were evaluated from each participant. The demographic variables evaluated in the study included gender, age, race and ethnicity. In addition, the following clinical data, including tumor characteristics were evaluated: tumor location, tumor stage (TNM stage), MSI, primary site of diagnosis, colon side, pathological group stage at diagnosis, mutation status of CRC related genes and medications used. Genetic similarity was modeled as a categorical variable (dichotomous for each genetic similairty population: less or equal than median genetic similarity levels).
Demographic, clinical and genomic characteristics among CRC cases were evaluated according to CRC status using Pearson's chi-square or Fisher's exact test, as appropriate for categorical variables. Overall, since a previous efforts [**11-12**] show no different of unadjusted and adjusted ORs, we present unadjusted ORs and its 95% confidence interval (CI) for the association of similarity and each tumor characteristic.

WES Comparison with Public Databases:
cBioPortal [**85**] for Cancer Genomics that is highly regarded resource for interactive exploration of multidimensional cancer genomics datasets was used to extract genomic data from publicly available datasets and comparisons were evaluated using Pearson's chi-square or Fisher's exact test, as appropriate.

Transcriptomics Comparison with Public Databases:
Transcriptomic data of 67 Non-Hispanic White was extracted from public databases projects TCGA-COAD and TCGA-READ projects. All individual demographic and clinical data from the 67 CRC tumors from Non-Hispanic White were matched to the same data from our 67 Hispanic and Latino CRC tumor samples. Differential Gene expression analyses were performed as mentioned before in the RNA sequencing analysis section.

Abbreviations
CRC: Colorectal Cancer; Hispanic and Latino: Hispanic/Latino; Non-Hispanic White: Non-Hispanic White;; MSI: Microsatellite instability; DGE: Differential Gene Expression; SCNA: Somatic Copy Number Alteration; SNP: Single nucleotide polymorphism; US: United States.

Acknowledgements
The authors would like to thank the Department of Integrative Translational Sciences at City of Hope (COH), the Cancer Control and Population Sciences program (P30CA033572) within the City of Hope Comprehensive Cancer Center , as well as, the NIH NCI grant project number U2CCA252971, the Genomic Characterization Unit and the Center for Optimizing Engagement of Hispanic Colorectal Cancer Patients in Cancer Genomic Characterization (COPECC) studies in the University of Southern California, Los Angeles. In addition, the authors would like to thank the patients, families, advocates, recruiters and all the personnel involved in the study for their contribution in this study.

Funding

This work was partially supported by the National Cancer Institute award number U2CCA252971 and the City of Hope Cancer Control and Population Sciences program award number P30CA033572.

REFERENCES

1.- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209-249. doi: 10.3322/caac.21660. Epub 2021 Feb 4. PMID: 33538338.

2.- Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. Nat Rev Gastroenterol Hepatol. 2019 Dec;16(12):713-732. doi: 10.1038/s41575-019-0189-8. Epub 2019 Aug 27. PMID: 31455888.

3.- World Health Organization (2024), International Agency for Research on Cancer. Global Cancer Observatory: Cancer Today. [Accessed on July 04, 2024. Available at: https://gco.iarc.who.int/media/globocan/factsheets/cancers/8-colon-fact-sheet.pdf and https://gco.iarc.who.int/media/globocan/factsheets/cancers/9-rectum-fact-sheet.pdf]

4.- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7-30.

5.- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018; 68:7–30. https://doi.org/10.3322/caac.21442.

6.- Pinheiro PS, Sherman RL, Trapido EJ, Fleming LE, Huang Y, Gomez-Marin O, Lee D. Cancer incidence in first generation U.S. Hispanics: Cubans, Mexicans, Puerto Ricans, and new Latinos. Cancer Epidemiol Biomarkers Prev. 2009 Aug;18(8):2162-9. doi: 10.1158/1055-9965.EPI-09-0329. PMID: 19661072.

7.- Heintzman JD, Ezekiel-Herrera DN, Quiñones AR, Lucas JA, Carroll JE, Gielbultowicz SH, Cottrell EC, Marino M. Disparities in Colorectal Cancer Screening in Latinos and Non-Hispanic Whites. Am J Prev Med. 2022 Feb;62(2):203-210. doi: 10.1016/j.amepre.2021.07.009. Epub 2021 Oct 11. PMID: 34649735.

8.- Ellis L, Canchola AJ, Spiegel D, Ladabaum U, Haile R, Gomez SL. Racial and Ethnic Disparities in Cancer Survival: The Contribution of Tumor, Sociodemographic, Institutional, and Neighborhood Characteristics. J Clin Oncol. 2018 Jan 1;36(1):25-33. doi: 10.1200/JCO.2017.74.2049. Epub 2017 Oct 16. PMID: 29035642; PMCID: PMC5756323.

9.- Miller KD, Ortiz AP, Pinheiro PS, Bandi P, Minihan A, Fuchs HE, Martinez Tyson D, Tortolero-Luna G, Fedewa SA, Jemal AM, Siegel RL. Cancer statistics for the US Hispanic/Latino population,

2021. CA Cancer J Clin. 2021 Nov;71(6):466-487. doi: 10.3322/caac.21695. Epub 2021 Sep 21. PMID: 34545941.

10.- Gonzalez-Pons M, Cruz-Correa M. Colorectal Cancer Disparities in Latinos: Genes vs. Environment. 2019 Dec 13. In: Ramirez AG, Trapido EJ, editors. Advancing the Science of Cancer in Latinos [Internet]. Cham (CH): Springer; 2020. Chapter 3. PMID: 34460206.

11.- Pérez-Mayoral J, Soto-Salgado M, Shah E, Kittles R, Stern MC, Olivera MI, Gonzalez-Pons M, Rodriguez-Quilichinni S, Torres M, Reyes JS, Tous L, López N, Chevere VC, Cruz-Correa M. Association of genetic ancestry with colorectal tumor location in Puerto Rican Latinos. Hum Genomics. 2019 Feb 20;13(1):12. doi: 10.1186/s40246-019-0196-4. PMID: 30786938; PMCID: PMC6383234.

12.- Perez-Mayoral J, Gonzalez-Pons M, Centeno-Girona H, Montes-Rodríguez IM, Soto-Salgado M, Suárez B, Rodríguez N, Colón G, Sevilla J, Jorge D, Llor X, Xicola RM, Toro DH, Tous-López L, Torres-Torres M, Reyes JS, López-Acevedo N, Goel A, Rodríguez-Quilichini S, Cruz-Correa M. Molecular and Sociodemographic Colorectal Cancer Disparities in Latinos Living in Puerto Rico. Genes (Basel). 2023 Apr 11;14(4):894. doi: 10.3390/genes14040894. PMID: 37107652; PMCID: PMC10138302.

13.- Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012 Jul 18;487(7407):330-7. doi: 10.1038/nature11252. PMID: 22810696; PMCID: PMC3401966.

14.- Fearon ER. Molecular genetics of colorectal cancer. Annu Rev Pathol. 2011;6:479-507. doi: 10.1146/annurev-pathol-011110-130235. PMID: 21090969.

15.- Li J, Ma X, Chakravarti D, Shalapour S, DePinho RA. Genetic and biological hallmarks of colorectal cancer. Genes Dev. 2021 Jun;35(11-12):787-820. doi: 10.1101/gad.348226.120. PMID: 34074695; PMCID: PMC8168558.

16.- Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21(11):1350-1356.

17.- Brooke Rhead, David M. Hein, Yannick Pouliot, Justin Guinney, View ORCID ProfileFrancisco M. De La Vega, Nina N. Sanford. Association of Genetic Ancestry with Molecular Tumor Profiles in Colorectal Cancer. MedRxiv (2023) doi: https://doi.org/10.1101/2023.07.12.23292571

18.- Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, Jass J, Lindblom A, Lynch HT, Peltomaki P, Ramsey SD, Rodriguez-Bigas MA, Vasen HF, Hawk ET, Barrett JC, Freedman AN, Srivastava S. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst. 2004 Feb 18;96(4):261-8. doi: 10.1093/jnci/djh034. PMID: 14970275; PMCID: PMC2933058.

19.- Heide T, Househam J, Cresswell GD, Spiteri I, Lynn C, Mossner M, Kimberley C, Fernandez-Mateos J, Chen B, Zapata L, James C, Barozzi I, Chkhaidze K, Nichol D, Gunasri V, Berner A, Schmidt M, Lakatos E, Baker AM, Costa H, Mitchinson M, Piazza R, Jansen M, Caravagna G, Ramazzotti D, Shibata D, Bridgewater J, Rodriguez-Justo M, Magnani L, Graham TA, Sottoriva A. The co-evolution of the genome and epigenome in colorectal cancer. Nature. 2022 Nov;611(7937):733-743. doi: 10.1038/s41586-022-05202-1. Epub 2022 Oct 26. PMID: 36289335; PMCID: PMC9684080.

20.- Mechanisms of KRAS Inhibitor Resistance Revealed. Cancer Discov. 2021 Jun;11(6):1311-1312. doi: 10.1158/2159-8290.CD-NB2021-0335. Epub 2021 Apr 14. PMID: 33853790.

21.- Bastide P, Darido C, Pannequin J, Kist R, Robine S, Marty-Double C, Bibeau F, Scherer G, Joubert D, Hollande F, Blache P, Jay P. Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. J Cell Biol. 2007 Aug 13;178(4):635-48. doi: 10.1083/jcb.200704152. PMID: 17698607; PMCID: PMC2064470.

22.- Mori-Akiyama Y, van den Born M, van Es JH, Hamilton SR, Adams HP, Zhang J, Clevers H, de Crombrugghe B. SOX9 is required for the differentiation of paneth cells in the intestinal epithelium. Gastroenterology. 2007 Aug;133(2):539-46. doi: 10.1053/j.gastro.2007.05.020. Epub 2007 May 21. PMID: 17681175.

23.- Fearon ER. Molecular genetics of colorectal cancer. Annu Rev Pathol. 2011;6:479-507. doi: 10.1146/annurev-pathol-011110-130235. PMID: 21090969.

24.- Tamas K, Walenkamp AM, de Vries EG, van Vugt MA, Beets-Tan RG, van Etten B, de Groot DJ, Hospers GA. Rectal and colon cancer: Not just a different anatomic site. Cancer Treat Rev. 2015 Sep;41(8):671-9. doi: 10.1016/j.ctrv.2015.06.007. Epub 2015 Jun 28. PMID: 26145760.

25.- Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhim, R., & Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biology, 12(4), R41. https://doi.org/10.1186/gb-2011-12-4-r41

26.- Mehrvarz Sarshekeh A, Advani S, Overman MJ, Manyam G, Kee BK, Fogelman DR, Dasari A, Raghav K, Vilar E, Manuel S, Shureiqi I, Wolff RA, Patel KP, Luthra R, Shaw K, Eng C, Maru DM, Routbort MJ, Meric-Bernstam F, Kopetz S. Association of SMAD4 mutation with patient demographics, tumor characteristics, and clinical outcomes in colorectal cancer. PLoS One. 2017 Mar 7;12(3):e0173345. doi: 10.1371/journal.pone.0173345. Erratum in: PLoS One. 2017 May 17;12(5):e0178275. doi: 10.1371/journal.pone.0178275. PMID: 28267766; PMCID: PMC5340382.

27.- Tanaka N, Mashima T, Mizutani A, Sato A, Aoyama A, Gong B, Yoshida H, Muramatsu Y, Nakata K, Matsuura M, Katayama R, Nagayama S, Fujita N, Sugimoto Y, Seimiya

H. *APC* Mutations as a Potential Biomarker for Sensitivity to Tankyrase Inhibitors in Colorectal Cancer. Mol Cancer Ther. 2017 Apr;16(4):752-762. doi: 10.1158/1535-7163.MCT-16-0578. Epub 2017 Feb 8. PMID: 28179481.

28.- Wheeler JM, Bodmer WF, Mortensen NJ. DNA mismatch repair genes and colorectal cancer. Gut. 2000 Jul;47(1):148-53. doi: 10.1136/gut.47.1.148. PMID: 10861278; PMCID: PMC1727951.

29.- Sailo BL, Banik K, Girisa S, Bordoloi D, Fan L, Halim CE, Wang H, Kumar AP, Zheng D, Mao X, Sethi G, Kunnumakkara AB. FBXW7 in Cancer: What Has Been Unraveled Thus Far? Cancers (Basel). 2019 Feb 19;11(2):246. doi: 10.3390/cancers11020246. PMID: 30791487; PMCID: PMC6406609.

30.- Kone AS, Ait Ssi S, Sahraoui S, Badou A. BTN3A: A Promising Immune Checkpoint for Cancer Prognosis and Treatment. Int J Mol Sci. 2022 Nov 3;23(21):13424. doi: 10.3390/ijms232113424. PMID: 36362212; PMCID: PMC9653866.

31.- Tieng FYF, Lee LH, Ab Mutalib NS. Deciphering colorectal cancer immune microenvironment transcriptional landscape on single cell resolution - A role for immunotherapy. Front Immunol. 2022 Aug 10;13:959705. doi: 10.3389/fimmu.2022.959705. PMID: 36032085; PMCID: PMC9399368.

32.- Toon CW, Chou A, Clarkson A, DeSilva K, Houang M, Chan JC, Sioson LL, Jankova L, Gill AJ. Immunohistochemistry for myc predicts survival in colorectal cancer. PLoS One. 2014 Feb 4;9(2):e87456. doi: 10.1371/journal.pone.0087456. PMID: 24503701; PMCID: PMC3913591.

33.- Janani B, Vijayakumar M, Priya K, Kim JH, Prabakaran DS, Shahid M, Al-Ghamdi S, Alsaidan M, Othman Bahakim N, Hassan Abdelzaher M, Ramesh T. EGFR-Based Targeted Therapy for Colorectal Cancer-Promises and Challenges. Vaccines (Basel). 2022 Mar 24;10(4):499. doi: 10.3390/vaccines10040499. PMID: 35455247; PMCID: PMC9030067.

34.- Ukkola I, Nummela P, Kero M, Tammio H, Tuominen J, Kairisto V, Kallajoki M, Haglund C, Peltomäki P, Kytölä S, Ristimäki A. Gene fusions and oncogenic mutations in MLH1 deficient and BRAFV600E wild-type colorectal cancers. Virchows Arch. 2022 Apr;480(4):807-817. doi: 10.1007/s00428-022-03302-x. Epub 2022 Mar 3. PMID: 35237889; PMCID: PMC9023403.

35.- Lipson D, Capelletti M, Yelensky R, Otto G, Parker A, Jarosz M, Curran JA, Balasubramanian S, Bloom T, Brennan KW, Donahue A, Downing SR, Frampton GM, Garcia L, Juhn F, Mitchell KC, White E, White J, Zwirko Z, Peretz T, Nechushtan H, Soussan-Gutman L, Kim J, Sasaki H, Kim HR, Park SI, Ercan D, Sheehan CE, Ross JS, Cronin MT, Jänne PA, Stephens PJ. Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. Nat Med. 2012 Feb 12;18(3):382-4. doi: 10.1038/nm.2673. PMID: 22327622; PMCID: PMC3916180.

36.- Kumar-Sinha C, Kalyana-Sundaram S, Chinnaiyan AM. Landscape of gene fusions in epithelial cancers: seq and ye shall find. Genome Med. 2015 Dec 18;7:129. doi: 10.1186/s13073-015-0252-1. PMID: 26684754; PMCID: PMC4683719.

37.- Wang J, Li R, Li J, Yi Y, Liu X, Chen J, Zhang H, Lu J, Li C, Wu H, Liang Z. Comprehensive analysis of oncogenic fusions in mismatch repair deficient colorectal carcinomas by sequential DNA and RNA next generation sequencing. J Transl Med. 2021 Oct 17;19(1):433. doi: 10.1186/s12967-021-03108-6. PMID: 34657620; PMCID: PMC8522100.

38.- Kloosterman WP, Coebergh van den Braak RRJ, Pieterse M, van Roosmalen MJ, Sieuwerts AM, Stangl C, Brunekreef R, Lalmahomed ZS, Ooft S, van Galen A, Smid M, Lefebvre A, Zwartkruis F, Martens JWM, Foekens JA, Biermann K, Koudijs MJ, Ijzermans JNM, Voest EE. A Systematic Analysis of Oncogenic Gene Fusions in Primary Colon Cancer. Cancer Res. 2017 Jul 15;77(14):3814-3822. doi: 10.1158/0008-5472.CAN-16-3563. Epub 2017 May 16. PMID: 28512242.

39.- Helsten, T. et al. (2016). The FGFR landscape in cancer: analysis of 4,853 tumors by next-generation sequencing. *Clinical Cancer Research*. 22(1), 259-267. doi:10.1158/1078-0432.CCR-14-3212.

40.- Babina IS, Turner NC. "Advances and challenges in targeting FGFR signalling in cancer." Nature Reviews Cancer. 2017.

41.- Katoh, M. (2019). Fibroblast growth factor receptors as treatment targets in clinical oncology. *Nature Reviews Clinical Oncology*. 16, 105-122. doi:10.1038/s41571-018-0117-y.

42.- Yoshida, Y. et al. (2023). Current Targeted Therapy for Metastatic Colorectal Cancer. *International Journal of Molecular Sciences*. 24(2), 1702. doi:10.3390/ijms24021702.

43.- Major MB, Camp ND, Berndt JD, Yi X, Goldenberg SJ, Hubbert C, Biechele TL, Gingras AC, Zheng N, Maccoss MJ, Angers S, Moon RT. Wilms tumor suppressor WTX negatively regulates WNT/beta-catenin signaling. Science. 2007 May 18;316(5827):1043-6. doi: 10.1126/science/1141515. PMID: 17510365.

44.- (32 & 147 repeated)- Jeong JH, Yun JW, Kim HY, Heo CY, Lee S. Investigation of cell signalings and therapeutic targets in PTPRK-RSPO3 fusion-positive colorectal cancer. PLoS One. 2022 Sep 21;17(9):e0274555. doi: 10.1371/journal.pone.0274555. PMID: 36129915; PMCID: PMC9491571.

45.- Kumar-Sinha C, Kalyana-Sundaram S, Chinnaiyan AM. Landscape of gene fusions in epithelial cancers: seq and ye shall find. Genome Med. 2015 Dec 18;7:129. doi: 10.1186/s13073-015-0252-1. PMID: 26684754; PMCID: PMC4683719.

46.- Helsten, T., Elkin, S., Arthur, E., Tomson, B. N., Carter, J., & Kurzrock, R. (2016). The FGFR landscape in cancer: Analysis of 4,853 tumors by next-generation sequencing. *Clinical Cancer Research, 22*(1), 259-267. doi:10.1158/1078-0432.CCR-14-3212.

47.- Seshagiri, S., Stawiski, E. W., Durinck, S., Modrusan, Z., Storm, E. E., Conboy, C. B., ... & de Sauvage, F. J. (2012). Recurrent R-spondin fusions in colon cancer. *Nature, 488*(7413), 660-664. doi:10.1038/nature11282.

48.- Massague J.,Blain,S.W.&Lo,R.S. TGFb signaling in growth control, cancer, and heritable disorders. Cell 103, 295–309 (2000).

49.- Fejerman L, Stern MC, Ziv E, John EM, Torres-Mejia G, Hines LM, et al. Genetic ancestry modifies the association between genetic risk variants and breast cancer risk among Hispanic and non-Hispanic white women. Carcinogenesis. 2013;34:1787–93. https://doi.org/10.1093/carcin/bgt110.

50.- Fejerman L, Hu D, Huntsman S, John EM, Stern MC, Haiman CA, et al. Genetic ancestry and risk of mortality among US Latinas with breast cancer. Cancer Res. 2013;73:7243–53.

51.- Al-Alem U, Rauscher G, Shah E, Batai K, Mahmoud A, Beisner E, et al. Association of genetic ancestry with breast cancer in ethnically diverse women from Chicago. PLoS One. 2014;9:e112916. https://doi.org/10.1371/ journal.pone.0112916.

52.- Bonilla C, Bertoni B, Hidalgo PC, Artagaveytia N, Ackermann E, Barreto I, et al. Breast cancer risk and genetic ancestry: a case-control study in Uruguay. BMC Womens Health. 2015;15:11. https://doi.org/10.1186/s12905-015-0171-8.

53.- Ricks-Santi LJ, Apprey V, Mason T, Wilson B, Abbas M, Hernandez W, et al. Identification of genetic risk associated with prostate cancer using ancestry informative markers. Prostate Cancer Prostatic Dis. 2012;15:359–64. https:// doi.org/10.1038/pcan.2012.19.

54.- Tian C, Hinds DA, Shigeta R, Kittles R, Ballinger DG, Seldin MF. A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. Am J Hum Genet. 2006;79:640–9. https://doi.org/10.1086/507954.

55.- Fejerman L, Chen GK, Eng C, Huntsman S, Hu D, Williams A, et al. Admixture mapping identifies a locus on 6q25 associated with breast cancer risk in US Latinas. Hum Mol Genet. 2012;21:1907–17. https://doi.org/10.1093/hmg/ddr617.

56.- Fejerman L, Stern MC, John EM, Torres-Mejia G, Hines LM, Wolff RK, et al. Interaction between common breast cancer susceptibility variants, genetic ancestry, and nongenetic risk factors in Hispanic women. Cancer Epidemiol Biomark Prev. 2015;24:1731–8. https://doi.org/10.1158/1055-9965.EPI-15-0392.

57.- Hernandez-Suarez G, Sanabria MC, Serrano M, Herran OF, Perez J, Plata JL, et al. Genetic ancestry is associated with colorectal adenomas and adenocarcinomas in Latino populations. Eur J Hum Genet. 2014;22:1208–16.

58.- Nagl, N. G. Jr, Zweitzig, D. R., Thimmapaya, B., Beck, G. R. Jr & Moran, E. The c-myc gene is a direct target of mammalian SWI/SNF-related complexes during differentiation-associated cell cycle arrest. Cancer Res. 66, 1289–1293 (2006).

59.- Polakis, P. (2000). Wnt signaling and cancer. Genes & Development, 14(15), 1837-1851. doi:10.1101/gad.14.15.1837

60.- Clevers, H., & Nusse, R. (2012). Wnt/β-catenin signaling and disease. Cell, 149(6), 1192-1205. doi:10.1016/j.cell.2012.05.012

61.- Chen, B. et al. Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. Nat. Chem. Biol. 5, 100–107 (2009).

62.- Ewan, K. et al. A useful approach to identify novel small-molecule inhibitors of Wnt-dependent transcription. Cancer Res. 70, 5963–5973 (2010).

63.- Sack, U. et al. S100A4-induced cell motility and metastasis is restricted by the Wnt/b-catenin pathway inhibitor calcimycin in colon cancer cells. Mol. Biol. Cell 22, 3344–3354 (2011).

64- 117.- Manojlovic Z, Christofferson A, Liang WS, et al. Comprehensive molecular profiling of 718 Multiple Myelomas reveals significant differences in mutation frequencies between African and European descent cases. PLoS Genet. 2017 Nov 22;13(11):e1007087. doi: 10.1371/journal.pgen.1007087. PMID: 29166413; PMCID: PMC5699827.

65- 118.- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754-60. doi: 10.1093/bioinformatics/btp324. Epub 2009 May 18. PMID: 19451168; PMCID: PMC2705234.

66- 119.- McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep;20(9):1297-303. doi: 10.1101/gr.107524.110. Epub 2010 Jul 19. PMID: 20644199; PMCID: PMC2928508.

67- 120.- Broad Institute. Picard toolkit (2023). [Accessed on July 28, 2023. Available at: https://broadinstitute.github.io/picard/].

68.- Kim S, Scheffler K, Halpern AL, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods. 2018 Aug;15(8):591-594. doi: 10.1038/s41592-018-0051-x. Epub 2018 Jul 16. PMID: 30013048.

69.- Benjamin D, Sato T, Cibulskis K, et al. Calling Somatic SNVs and Indels with Mutect2. bioRxiv 861054; doi: https://doi.org/10.1101/861054.

70.- Robinson JT, Thorvaldsdóttir H, Wenger AM, et al. Variant Review with the Integrative Genomics Viewer. Cancer Res. 2017 Nov 1;77(21):e31-e34. doi: 10.1158/0008-5472.CAN-17-0337. PMID: 29092934; PMCID: PMC5678989.

71.- McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016 Jun 6;17(1):122. doi: 10.1186/s13059-016-0974-4. PMID: 27268795; PMCID: PMC4893825.

72.- Mayakonda A, Lin DC, Assenov Y, et al. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018 Nov;28(11):1747-1756. doi: 10.1101/gr.239244.118. Epub 2018 Oct 19. PMID: 30341162; PMCID: PMC6211645.

73.- 1000 Genomes Project Consortium; Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. Nature. 2015 Oct 1;526(7571):68-74. doi: 10.1038/nature15393. PMID: 26432245; PMCID: PMC4750478.

74.- Danecek P, Auton A, Abecasis G, et al; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. Bioinformatics. 2011 Aug 1;27(15):2156-8. doi: 10.1093/bioinformatics/btr330. Epub 2011 Jun 7. PMID: 21653522; PMCID: PMC3137218.

75.- Cingolani P, Patel VM, Coon M, et al. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front Genet. 2012 Mar 15;3:35. doi: 10.3389/fgene.2012.00035. PMID: 22435069; PMCID: PMC3304048..

76.- Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015 Feb 25;4:7. doi: 10.1186/s13742-015-0047-8. PMID: 25722852; PMCID: PMC4342193.

77.- Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 2003 Aug;164(4):1567-87. doi: 10.1093/genetics/164.4.1567. PMID: 12930761; PMCID: PMC1462648.

78.- Baran Y, Pasaniuc B, Sankararaman S, et al. Fast and accurate inference of local ancestry in Latino populations. Bioinformatics. 2012 May 15;28(10):1359-67. doi: 10.1093/bioinformatics/bts144. Epub 2012 Apr 11. PMID: 22495753; PMCID: PMC3348558.

79.- Wickham, Hadley. Ggplot2: Elegant Graphics for Data Analysis. 2nd ed. Springer International Publishing, 2016. https://doi.org/10.1007/978-3-319-24277-4.

80.- Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. 2019 May 7;47(8):e47. doi: 10.1093/nar/gkz114. PMID: 30783653; PMCID: PMC6486549.

81.- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017 Apr;14(4):417-419. doi: 10.1038/nmeth.4197. Epub 2017 Mar 6. PMID: 28263959; PMCID: PMC5600148.

82.- Nowicka M, Robinson MD. DRIMSeq: a Dirichlet-multinomial framework for multivariate count outcomes in genomics. F1000Res. 2016 Jun 13;5:1356. doi: 10.12688/f1000research.8900.2. PMID: 28105305; PMCID: PMC5200948.

83.- Long MD, Jacobi JJ, Singh PK, Llimos G, Wani SA, Rowsam AM, Rosario SR, Hoogstraat M, Linder S, Kirk J, Affronti HC, Bergman A, Zwart W, Campbell MJ, Smiraglia DJ. Reduced NCOR2 expression accelerates androgen deprivation therapy failure in prostate cancer. Cell Rep. 2021 Dec 14;37(11):110109. doi: 10.1016/j.celrep.2021.110109. PMID: 34910907; PMCID: PMC8889623.

84.- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15-21. doi: 10.1093/bioinformatics/bts635. Epub 2012 Oct 25. PMID: 23104886; PMCID: PMC3530905.

85.- Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012 May;2(5):401-4. doi: 10.1158/2159-8290.CD-12-0095. Erratum in: Cancer Discov. 2012 Oct;2(10):960. PMID: 22588877; PMCID: PMC3956037.

86.- Loevenich LP, Tschurtschenthaler M, Rokavec M, Silva MG, Jesinghaus M, Kirchner T, Klauschen F, Saur D, Neumann J, Hermeking H, Jung P. SMAD4 Loss Induces c-MYC-Mediated NLE1 Upregulation to Support Protein Biosynthesis, Colorectal Cancer Growth, and Metastasis. Cancer Res. 2022 Dec 16;82(24):4604-4623. doi: 10.1158/0008-5472.CAN-22-1247. PMID: 36219392; PMCID: PMC9755967.

87.- Szeglin BC, Wu C, Marco MR, Park HS, Zhang Z, Zhang B, Garcia-Aguilar J, Beauchamp RD, Chen XS, Smith JJ. A SMAD4-modulated gene profile predicts disease-free survival in stage II and III colorectal cancer. Cancer Rep (Hoboken). 2022 Jan;5(1):e1423. doi: 10.1002/cnr2.1423. Epub 2021 Jun 10. PMID: 34114372; PMCID: PMC8789617.

88.- Testa U, Pelosi E, Castelli G. Colorectal cancer: genetic abnormalities, tumor progression, tumor heterogeneity, clonal evolution and tumor-initiating cells. Med Sci (Basel). 2018 Apr 13;6(2):31. doi: 10.3390/medsci6020031. PMID: 29652830; PMCID: PMC6024750.

89.- Vaccaro CA, López-Kostner F, Adriana DV, Palmero EI, Rossi BM, Antelo M, Solano A, Carraro DM, Forones NM, Bohorquez M, Lino-Silva LS, Buleje J, Spirandelli F, Abe-Sandes K, Nascimento I, Sullcahuaman Y, Sarroca C, Gonzalez ML, Herrando AI, Alvarez K, Neffa F, Galvão HC, Esperon P, Golubicki M, Cisterna D, Cardoso FC, Torrezan GT, Junior SA, Pimenta CAM, da Cruz Formiga MN, Santos E, Sá CU, Oliveira EP, Fujita R, Spirandelli E, Jimenez G, Guindalini RSC, de Azevedo RGMV, Bueno LSM, Dos Santos Nogueira ST, Loarte MT, Padron J, Del Carmen Castro-Mujica M, Del Monte JS, Caballero C, Peña CMM, Pinto J, Barletta-Carrillo C, Melva GA, Piñero T, Beltran PM, Ashton-Prolla P, Rodriguez Y, Quispe R, Rossi NT, Martin C, Chialina S, Kalfayan PG, Bazo-Alvarez JC, Cañete AR, Dominguez-Barrera C, Nuñez L, Da Silva SD, Balavarca Y, Wernhoff P, Plazzer JP, Møller P, Hovig E, Dominguez-Valentin M; GETH. From colorectal cancer pattern to the characterization of individuals at risk: Picture for genetic research in Latin America. Int J Cancer. 2019 Jul 15;145(2):318-326. doi: 10.1002/ijc.31920. Epub 2018 Dec 5. PMID: 30303536; PMCID: PMC6587543.

90.- Johnson RM, Qu X, Lin CF, Huw LY, Venkatanarayan A, Sokol E, Ou FS, Ihuegbu N, Zill OA, Kabbarah O, Wang L, Bourgon R, de Sousa E Melo F, Bolen C, Daemen A, Venook AP, Innocenti F, Lenz HJ, Bais C. ARID1A mutations confer intrinsic and acquired resistance to cetuximab treatment in colorectal cancer. Nat Commun. 2022 Sep 19;13(1):5478. doi: 10.1038/s41467-022-33172-5. PMID: 36117191; PMCID: PMC9482920.

Table 1. Patient Demographics, Clinical Characteristics, and Genetic Similarity Proportions. This table provides an overview of patient demographics (with race and ethnicity self-reported), clinical characteristics, and genetic similarity proportions relative to the 1000 Genomes Project reference populations. These genetic similarity proportions are categorized as 1KG-PEL-like (Peruvian-in-Lima-like), 1KG-EUR-like (European-like), 1KG-AFR-like (African-like), 1KG-EAS-like (East Asian-like), and 1KG-SAS-like (South Asian-like). The genetic similarity proportions offer insights into the continental origins of the patients' genomes, enhancing our understanding of their genetic background.

Characteristic	CRC Moonshot (n=67) *n (%)*
Age at Diagnosis	
< 50 years	30 (44.8%)
>50 years	37 (55.2%)
Gender	
Male	36 (53.7%)
Female	31 (46.3%)
MSI Status	
MSS	54 (80.6%)
MSI	13 (19.4%)
Tumor Mutational Burden	
Non-Hypermutated	51 (76.1%)
Hypermutated	16 (23.9%)
Tumor Location	
Colon	42 (62.7%)
Rectum	25 (37.3%)
Pathological Stage	
Early (Stage I or II)	19 (28.4%)
Late (Stage III or IV)	29 (43.3%)
No TNM Applicable/Unknown	19 (28.4%)
1KG-PEL-like Similarity (mean ± SD)	0.58 ± 0.14
1KG-EUR-like Similarity Ancestry (mean ± SD)	0.36 ± 0.13
1KG-AFR-like Similarity (mean ± SD)	0.04 ± 0.04
1KG-EAS-like Similarity (mean ± SD)	0.01 ± 0.02
1KG-SAS-like SImilarity (mean ± SD)	0.01 ± 0.02

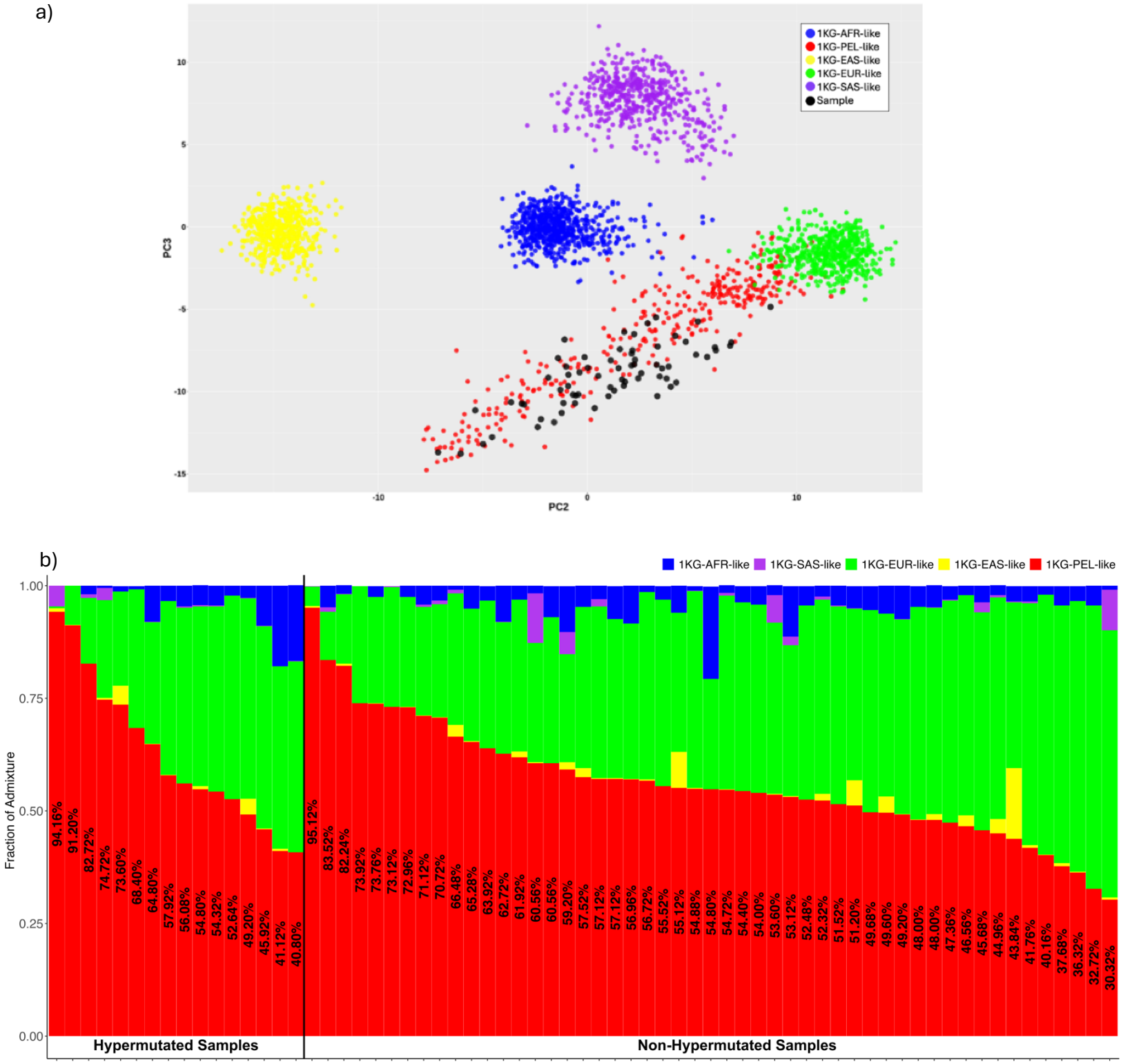


Figure 1: Genetic Similarity and Chromosome-Level Patterns. a) Genetic similarity principal component analysis (PCA) plot. This PCA plot illustrates the genetic similarity of our Hispanic/Latino colorectal cancer cohort of 67 individuals, represented by black points, in the context of five super populations: 1000 Genomes Project African-like (1KG-AFR-like, blue), 1000 Genomes Project Peruvian-in-Lima-like (1KG-PEL-like, red), 1000 Genomes Project East-Asian-like (1KG-EAS-like, yellow), 1000 Genomes Project European-like (1KG-EUR-like, green), and 1000 Genomes Project South-Asian-like (1KG-SAS-like, purple). The plot displays the distribution of these samples along two principal components, PC2 and PC3, which capture the majority of the variance in the genetic data. This visualization helps to contextualize the genetic similarity composition of our cohort relative to these major global populations. **b)** This figure illustrates the frequencies in each tumor sample from a cohort of 67 Hispanic/Latino (H/L) colorectal cancer (CRC) patients. The genomes are categorized into five super populations, same as section a. Samples are stratified into hypermutated and non-hypermutated groups, with the proportion of 1KG-PEL-like similarity included.

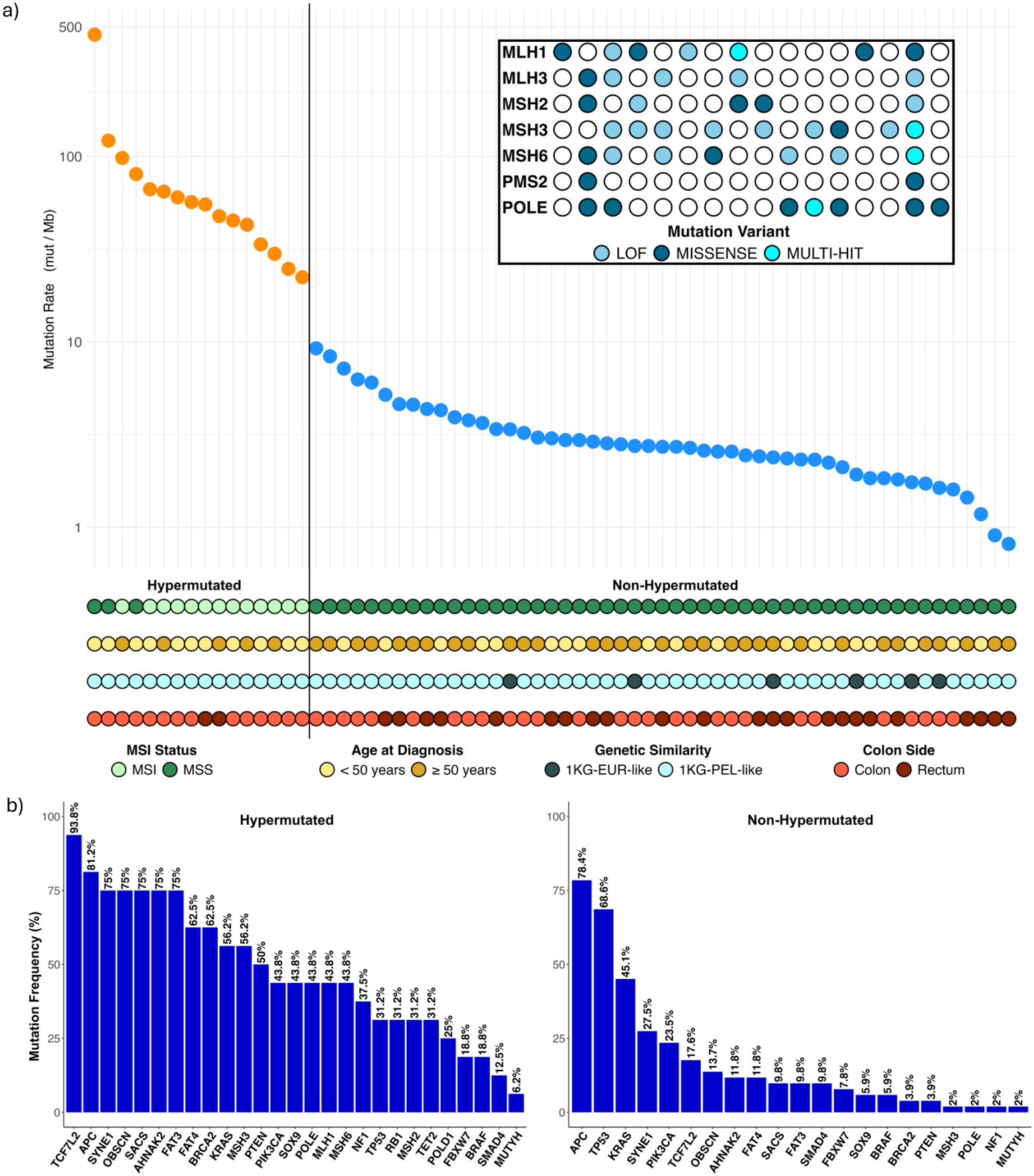


Figure 2: Mutation Frequencies in CRC Tumors from H/L Patients. a) This panel shows the mutation frequencies in each of the tumor samples from 67 Hispanic/Latino colorectal cancer patients. The samples are categorized as hypermutated or non-hypermutated. The color codes represent different attributes: light green for MSI, dark green for MSS, light yellow for age at diagnosis < 50 years, brown for age at diagnosis > 50 years, light blue for highest 1KG-PEL-like similarity proportions, dark blue for highest 1KG-EUR-like similarity proportions, light red for colon tumors, and dark red for rectum tumors. The inset highlights mutations in mismatch-repair genes and POLE among the hypermutated samples, with the sample order consistent with the main graph. **b)** This panel depicts the significantly mutated genes in hypermutated and non-hypermutated tumors.

Table 2. Mutation Frequencies Across CRC Studies. This table includes gene names, mutation frequencies from the current Moonshot study (N=67), and frequencies from external studies: MSK-NatCommun (N= 148), MSK-JNCI (N= 1513), DFCI-CellReports (N= 619), TCGA-Nature (N= 224), TCGA-PanCancerAtlas (N= 534), MSK-CancerDiscovery (N= 22), MSK-Gastroenterology (N= 471), MSK-JCO-PrecisOncol (N= 47). The included p-values demonstrate the significance levels when comparing frequencies with our study.

	Moonshot Study (N = 67)	MSK-NatCommun (N = 148)		MSK-JNCI (N = 1513)		DFCI-CellReports (N = 619)		TCGA-Nature (N = 224)	
Gene	Mutated *n (%)*	Mutated *n (%)*	p-value	Mutated *n (%)*	p-value	Mutated *n (%)*	p-value	Mutated *n (%)*	p-value
APC	54 (80.60%)	35 (23.60%)	**1.333E-14**	1145 (75.50%)	0.4227	361 (58.30%)	**6.46E-04**	168 (75.00%)	0.4346
TP53	40 (59.70%)	131 (88.50%)	**3.05E-06**	1107 (73.00%)	**0.0245**	320 (51.70%)	0.2637	121 (54.00%)	0.4959
KRAS	34 (50.75%)	41 (27.70%)	**0.0018**	666 (43.90%)	0.3303	173 (27.90%)	**1.98E-04**	94 (42.00%)	0.2583
SYNE1	26 (38.81%)	0 (NA)	1	0 (NA)	1	167 (27.00%)	0.0572	47 (21.00%)	**0.0052**
TCF7L2	25 (37.31%)	0 (NA)	1	196 (14.20%)	**6.725E-07**	44 (7.10%)	**3.089E-14**	27 (12.10%)	**5.28E-06**
OBSCN	19 (28.36%)	0 (NA)	1	0 (NA)	1	116 (18.70%)	0.0856	14 (6.30%)	**1.69E-06**
PIK3CA	19 (28.36%)	9 (6.10%)	**1.90E-05**	298 (19.70%)	0.1128	132 (21.30%)	0.2441	45 (20.10%)	0.2057
SACS	18 (26.87%)	0 (NA)	1	0 (NA)	1	67 (10.80%)	**0.0003**	22 (9.80%)	**8.01E-04**
AHNAK2	18 (26.87%)	0 (NA)	1	0 (NA)	1	106 (17.10%)	0.0717	21 (9.40%)	**4.96E-04**
FAT3	17 (25.37%)	9 (6.10%)	**1.49E-04**	0 (NA)	1	90 (14.50%)	**0.0320**	16 (7.10%)	**9.26E-05**
FAT4	16 (23.88%)	0 (NA)	1	0 (NA)	1	109 (17.60%)	0.2728	39 (17.40%)	0.3130
BRCA2	12 (17.91%)	5 (3.40%)	**7.13E-04**	83 (5.50%)	**8.45E-05**	42 (6.80%)	**0.0029**	10 (4.50%)	**7.01E-04**
MSH3	10 (14.93%)	2 (2.20%)	**0.0042**	28 (4.70%)	**0.0018**	20 (3.20%)	**3.60E-05**	15 (6.70%)	**0.0460**
SOX9	10 (14.93%)	4 (2.70%)	**0.0016**	154 (10.20%)	0.2945	62 (10.00%)	0.3004	37 (4.50%)	0.9032
PTEN	10 (14.93%)	4 (2.70%)	**0.0016**	88 (5.80%)	**0.0056**	51 (8.20%)	0.1095	8 (3.60%)	**1.96E-03**
POLE	8 (11.94%)	4 (2.70%)	**0.0101**	85 (5.60%)	0.0585	47 (7.60%)	0.3135	15 (6.70%)	0.2552
FBXW7	7 (10.45%)	20 (13.50%)	0.6847	205 (13.50%)	0.5893	86 (13.90%)	0.5520	37 (16.50%)	0.3066
NF1	7 (10.45%)	9 (6.10%)	0.3957	76 (5.00%)	0.0943	45 (7.30%)	0.4898	8 (3.60%)	0.0551
MLH1	7 (10.45%)	1 (0.70%)	**0.0014**	41 (2.70%)	**0.0011**	24 (3.90%)	**0.0316**	6 (2.70%)	**0.0181**
MSH6	7 (10.45%)	3 (2.00%)	**0.0114**	59 (3.90%)	**0.0206**	27 (4.40%)	0.0596	15 (6.70%)	0.4498
SMAD4	7 (10.45%)	19 (12.80%)	0.7856	238 (15.70%)	0.3219	73 (11.80%)	0.9001	26 (11.60%)	0.9657
BRAF	6 (8.96%)	6 (4.10%)	0.2588	135 (8.90%)	1	127 (20.50%)	**0.0348**	21 (9.40%)	1
RB1	5 (7.46%)	1 (0.70%)	**0.0120**	36 (2.40%)	**0.0298**	22 (3.60%)	0.2179	5 (2.20%)	0.0930
MSH2	5 (7.46%)	2 (1.40%)	**0.0316**	38 (2.50%)	**0.0396**	12 (1.90%)	**0.0188**	5 (2.20%)	0.0930
TET2	5 (7.46%)	3 (2.00%)	0.1113	53 (3.50%)	0.1742	11 (1.80%)	**0.0123**	4 (1.80%)	**0.0328**
POLD1	4 (5.97%)	3 (2.10%)	0.2098	45 (3.30%)	0.2818	34 (5.50%)	0.7805	5 (2.20%)	0.2188
MUTYH	2 (2.99%)	3 (2.00%)	0.6480	16 (1.10%)	0.1751	8 (1.30%)	0.2545	2 (0.90%)	0.2280

	Moonshot Study (N = 67)	TCGA-PanCancerAtlas (N = 534)		MSK-CancerDiscovery (N = 22)		MSK-Gastroenterology (N = 471)		MSK-JCO-PrecisOncol (N = 47)	
Gene	Mutated n (%)	Mutated n (%)	p-value	Mutated n (%)	p-value	Mutated n (%)	p-value	Mutated n (%)	p-value
APC	54 (80.60%)	387 (72.50%)	0.2035	21 (95.50%)	0.174	325 (69.00%)	0.07137	39 (83.00%)	0.9382
TP53	40 (59.70%)	314 (58.80%)	0.9925	21 (95.50%)	**0.00124**	345 (73.20%)	**0.03115**	20 (42.60%)	0.1064
KRAS	34 (50.75%)	218 (40.80%)	0.1556	22 (100.00%)	**6.33E-06**	238 (50.50%)	1	45 (95.70%)	**8.217E-08**
SYNE1	26 (38.81%)	144 (27.00%)	0.05952	0 (NA)	1	0 (NA)	1	0 (NA)	1
TCF7L2	25 (37.31%)	58 (10.90%)	**1.018E-08**	11 (50.00%)	0.1797	36 (9.20%)	**1.251E-09**	17 (37.80%)	1
OBSCN	19 (28.36%)	99 (18.50%)	0.08115	0 (NA)	1	0 (NA)	1	0 (NA)	1
PIK3CA	19 (28.36%)	147 (27.50%)	1	9 (40.90%)	0.4035	88 (18.70%)	0.09051	17 (36.20%)	0.4974
SACS	18 (26.87%)	67 (12.50%)	**0.002841**	0 (NA)	1	0 (NA)	1	0 (NA)	1
AHNAK2	18 (26.87%)	65 (12.20%)	**0.001947**	0 (NA)	1	0 (NA)	1	0 (NA)	1
FAT3	17 (25.37%)	81 (15.20%)	**0.05048**	0 (NA)	1	0 (NA)	1	0 (NA)	1
FAT4	16 (23.88%)	124 (23.20%)	**0.05296**	0 (NA)	1	0 (NA)	1	0 (NA)	1
BRCA2	12 (17.91%)	38 (7.10%)	**0.005419**	3 (13.60%)	1	21 (4.50%)	**5.78E-05**	7 (14.90%)	0.8004
MSH3	10 (14.93%)	11 (2.10%)	**2.02E-05**	3 (13.60%)	1	6 (3.90%)	**0.009045**	5 (12.50%)	0.9507
SOX9	10 (14.93%)	64 (12.00%)	0.6219	8 (36.40%)	0.06202	39 (8.30%)	0.1231	11 (23.40%)	0.3659
PTEN	10 (14.93%)	34 (6.40%)	**0.02224**	4 (18.20%)	0.7407	23 (4.90%)	**0.003355**	10 (21.30%)	0.4558
POLE	8 (11.94%)	36 (6.70%)	0.1967	7 (31.80%)	0.06683	17 (3.60%)	**0.006509**	10 (21.30%)	0.278
FBXW7	7 (10.45%)	90 (16.90%)	0.2431	8 (36.40%)	**0.0128**	51 (10.80%)	1	13 (27.70%)	**0.03331**
NF1	7 (10.45%)	31 (5.80%)	0.228	5 (22.70%)	0.2698	24 (5.10%)	0.1392	6 (12.80%)	0.933
MLH1	7 (10.45%)	22 (4.10%)	**0.04817**	10 (45.50%)	**9.28E-04**	9 (1.90%)	**5.31E-04**	2 (4.30%)	0.3032
MSH6	7 (10.45%)	24 (4.50%)	0.07446	8 (36.40%)	**0.0128**	6 (1.30%)	**3.32E-05**	8 (17.00%)	0.4589
SMAD4	7 (10.45%)	68 (12.70%)	0.7356	14 (63.60%)	**1.52E-06**	69 (14.60%)	0.4614	11 (23.40%)	0.1081
BRAF	6 (8.96%)	62 (11.60%)	0.6584	7 (31.80%)	**0.008478**	58 (12.30%)	0.5532	4 (8.50%)	1
RB1	5 (7.46%)	14 (2.60%)	0.07767	8 (36.40%)	**0.000785**	7 (1.50%)	**0.007873**	3 (6.40%)	1
MSH2	5 (7.46%)	21 (3.90%)	0.3076	5 (22.70%)	0.1146	4 (0.80%)	**0.002197**	3 (6.40%)	1
TET2	5 (7.46%)	23 (4.30%)	0.3966	12 (54.50%)	**5.07E-06**	13 (2.80%)	0.1011	3 (6.40%)	1
POLD1	4 (5.97%)	29 (5.40%)	0.7777	5 (22.70%)	**0.03801**	7 (1.80%)	0.06152	5 (11.10%)	0.4804
MUTYH	2 (2.99%)	7 (1.30%)	0.2648	3 (13.60%)	0.09447	3 (0.60%)	0.1192	1 (2.10%)	1

Table 3: Mutation Frequencies in Hispanic/Latino and Non-Hispanic White Cases. This table compares mutation frequencies in Hispanic/Latino (H/L) and Non-Hispanic White (NHW) cases across three datasets: our Moonshot study, the AACR Project GENIE dataset, and the TCGA-PanCancer project. It lists gene names along with their respective mutation frequencies as percentages for each racial group. The final column presents p-values, indicating the statistical significance of the observed differences in mutation frequencies between H/L and NHW patients.

Gene	H/L Mutations (N = 140)	NHW Mutations (N = 1487)	p-value
APC	112 (80.00%)	1075 (72.29%)	0.06245
TP53	89 (63.57%)	1038 (69.80%)	0.152
KRAS	69 (49.29%)	624 (41.96%)	0.1128
SYNE1	28 (20.00%)	80 (5.38%)	**1.01E-10**
TCF7L2	32 (22.86%)	156 (10.49%)	**2.26E-05**
OBSCN	19 (13.57%)	59 (3.97%)	**1.07E-06**
PIK3CA	46 (32.86%)	323 (21.72%)	**0.003703**
SACS	19 (13.57%)	30 (2.02%)	**1.49E-13**
AHNAK2	19 (13.57%)	33 (2.22%)	**1.80E-12**
FAT3	17 (12.14%)	44 (2.96%)	**1.64E-07**
FAT4	16 (11.43%)	67 (4.51%)	**0.0007846**
BRCA2	18 (12.86%)	116 (7.80%)	0.0549
MSH3	12 (8.57%)	25 (1.68%)	**8.16E-07**
SOX9	18 (12.86%)	170 (11.43%)	0.7145
PTEN	15 (10.71%)	95 (6.39%)	0.07627
POLE	14 (10.00%)	83 (5.58%)	0.05435
FBXW7	17 (12.14%)	197 (13.25%)	0.811
NF1	9 (6.43%)	84 (5.65%)	0.8497
MLH1	9 (6.43%)	52 (3.50%)	0.1303
MSH6	9 (6.43%)	70 (4.71%)	0.4838
SMAD4	19 (13.57%)	216 (14.53%)	0.8561
BRAF	13 (9.29%)	195 (13.11%)	0.2443
RB1	8 (5.71%)	31 (2.08%)	**0.01661**
MSH2	8 (5.71%)	47 (3.16%)	0.1758
TET2	9 (6.43%)	67 (4.51%)	0.4115
POLD1	7 (5.00%)	71 (4.77%)	1
MUTYH	3 (2.14%)	22 (1.48%)	0.8021

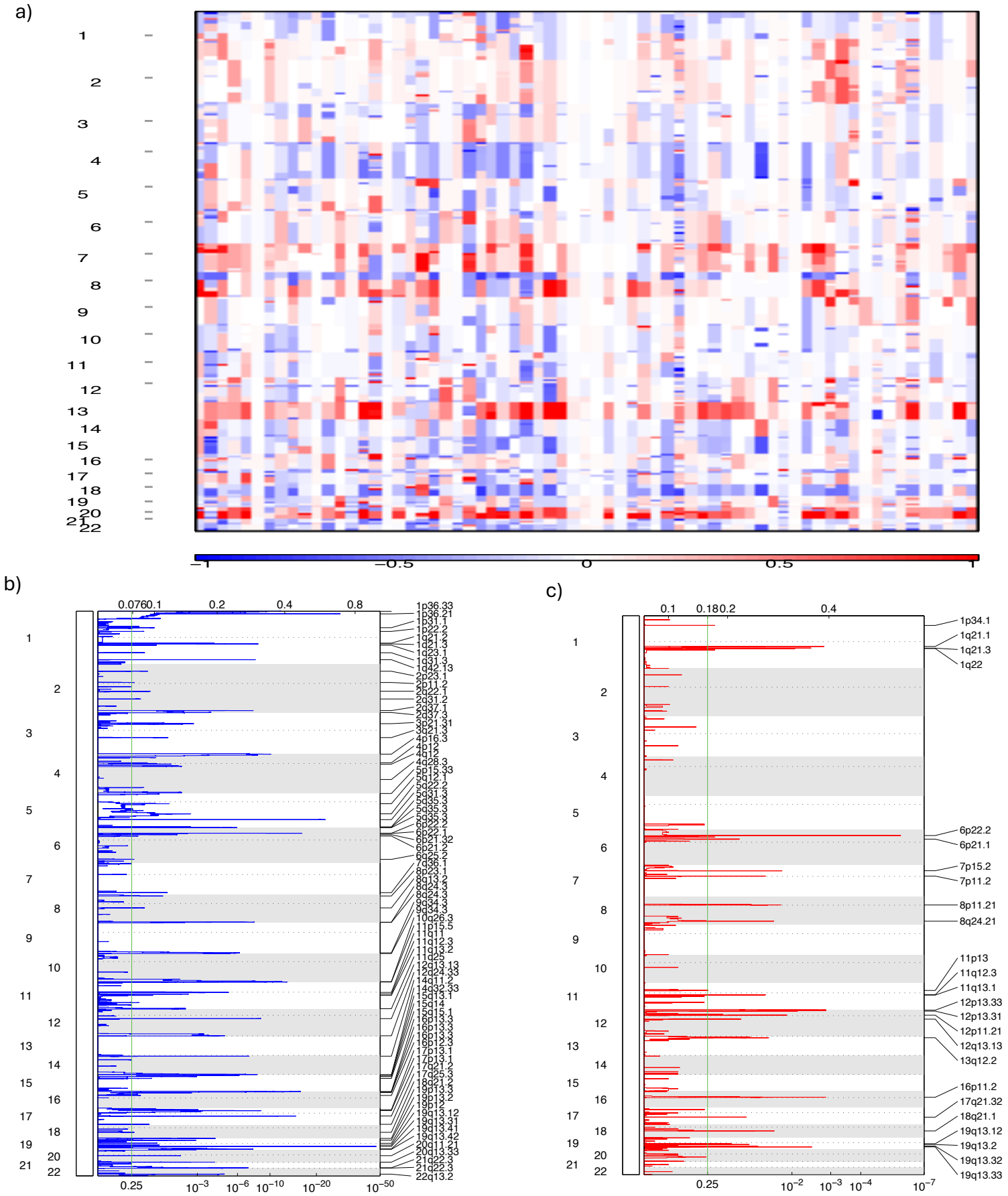
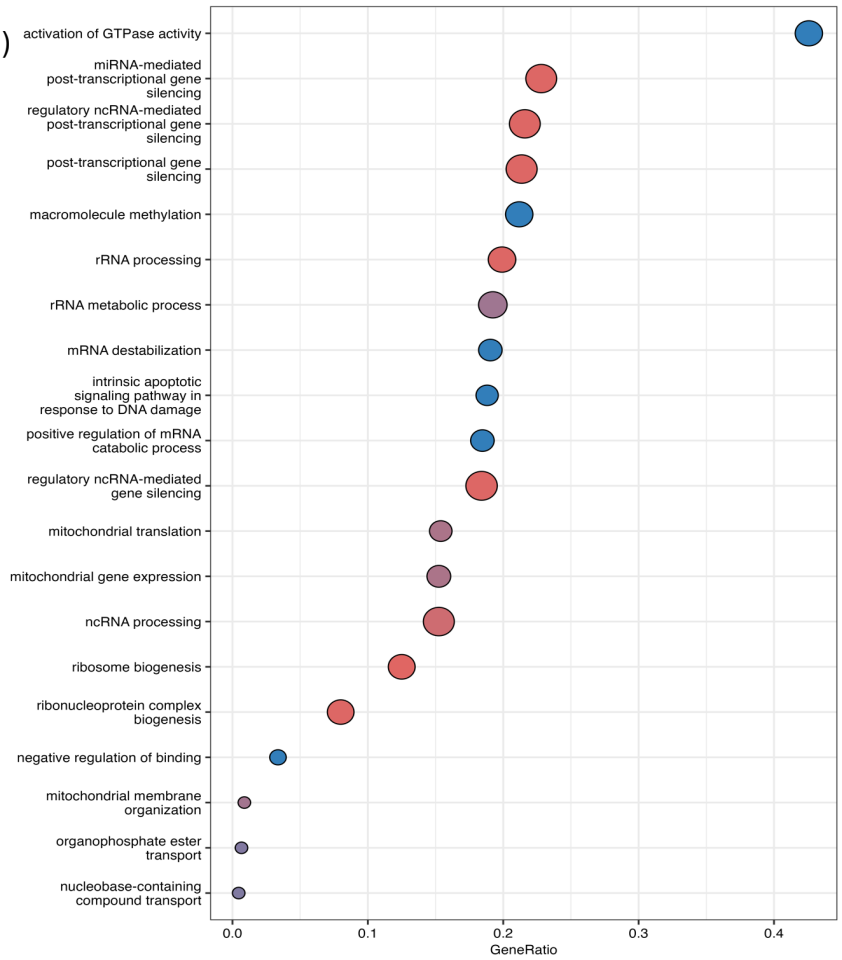


Figure 3: Somatic Copy Number Alterations (SCNAs) Analysis of Genomic Changes in 67 Hispanic/Latino Colorectal Cancer Patients. a) Heatmap of SCNAs: This panel displays SCNAs at both chromosomal and sub-chromosomal levels across 67 tumor samples from Hispanic/Latino colorectal cancer patients. Deletions (losses) are indicated in blue, while insertions (gains) are shown in red. **b)** Focal Deletions: This figure highlights the focal deletions identified across all 67 tumor samples. **c)** Focal Amplifications: This figure presents the focal amplifications observed in the same set of tumor samples.

a)

Gene	logFC	logCPM	LR	PValue	FDR
PIK3CA	1.76	5.20	56.98	4.407E-14	4.62E-13
MUTYH	-2.39	3.12	46.65	8.487E-12	7.45E-11
PMS2P3	-1.99	3.24	40.26	2.226E-10	1.70E-09
POLE	-1.78	6.59	31.17	2.362E-08	1.41E-07
PTEN	1.33	6.88	29.24	6.385E-08	3.60E-07
SYNE1	2.40	6.62	21.12	4.313E-06	1.82E-05
POLD1	-1.73	5.27	19.64	9.354E-06	3.72E-05
SMAD4	1.12	6.16	19.35	1.091E-05	4.29E-05
HFM1	3.31	-0.68	19.33	1.099E-05	4.32E-05
FAT3	3.01	2.05	18.87	1.402E-05	5.40E-05
TP53	-1.74	6.96	14.67	1.279E-04	4.10E-04
TCF7L2	-1.06	6.78	14.55	1.368E-04	4.37E-04
RB1	1.13	6.19	13.89	1.942E-04	6.01E-04
KRAS	1.08	6.14	12.33	4.467E-04	1.28E-03
TET2	0.86	5.91	10.08	0.001500	0.0038
MSH5	-1.31	3.90	9.02	0.002676	0.0064
BRAF	0.58	6.16	8.26	0.004046	0.0093
APC	0.83	5.56	7.80	0.005228	0.0117
PMS1	0.75	4.45	7.75	0.005364	0.0120
FAT4	1.26	5.52	7.75	0.005365	0.0120
OBSCN	-1.14	5.35	5.68	0.017131	0.0335
SACS	0.82	5.03	4.85	0.027708	0.0511

c)



b)

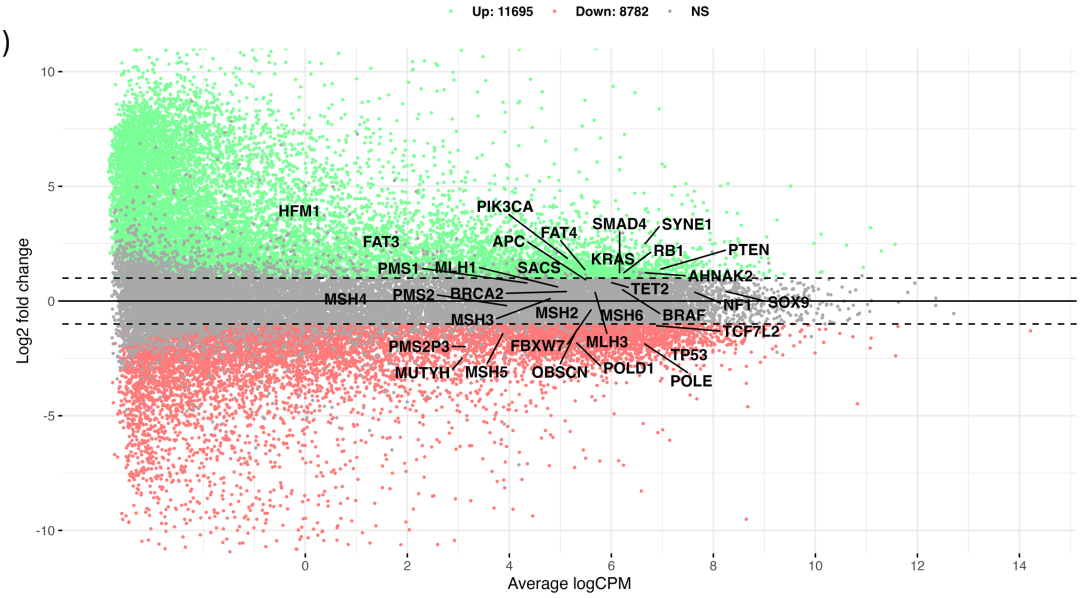
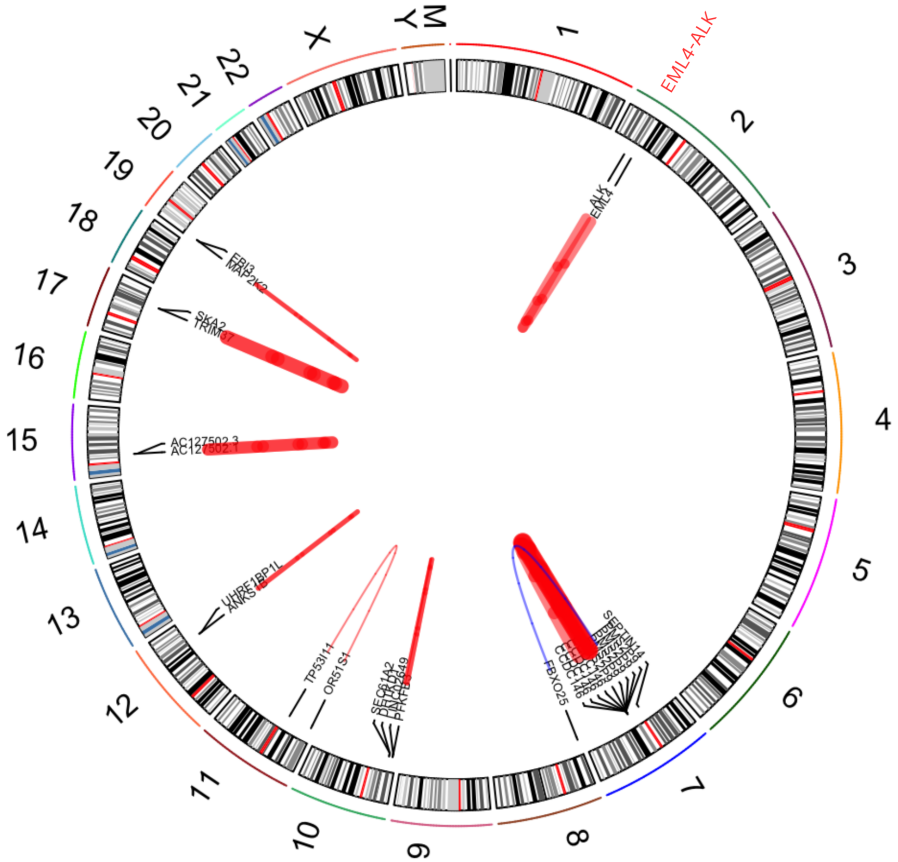


Figure 4: Differential gene expression (DGE) analysis among CRC tumors from 67 Hispanic/Latino patients and 67 Non-Hispanic White patients. DGE analysis identified unique gene expression patterns in our Hispanic/Latino (H/L) CRC cohort compared to Non-Hispanic White (NHW) CRC cases. The NHW cases were individually matched to our CRC cohort using demographic, clinical, and genomic data from two publicly available databases: TCGA-COAD and TCGA-READ. **a)** Differentially expressed CRC-related genes: This table shows the genes that are differentially expressed when comparing H/L and NHW CRC cases. **b)** Mean-Average (MA) plot: This scatter plot shows the upregulated and downregulated CRC-related genes according to the log2 fold change. **c)** Pathway analysis: This graph presents the distinct cellular pathways identified through pathway analysis by comparing H/L and NHW CRC cases. The graph displays the names of the pathways, adjusted p-values, and the number of genes altered in each pathway.

Hypermutated



Non-Hypermutated

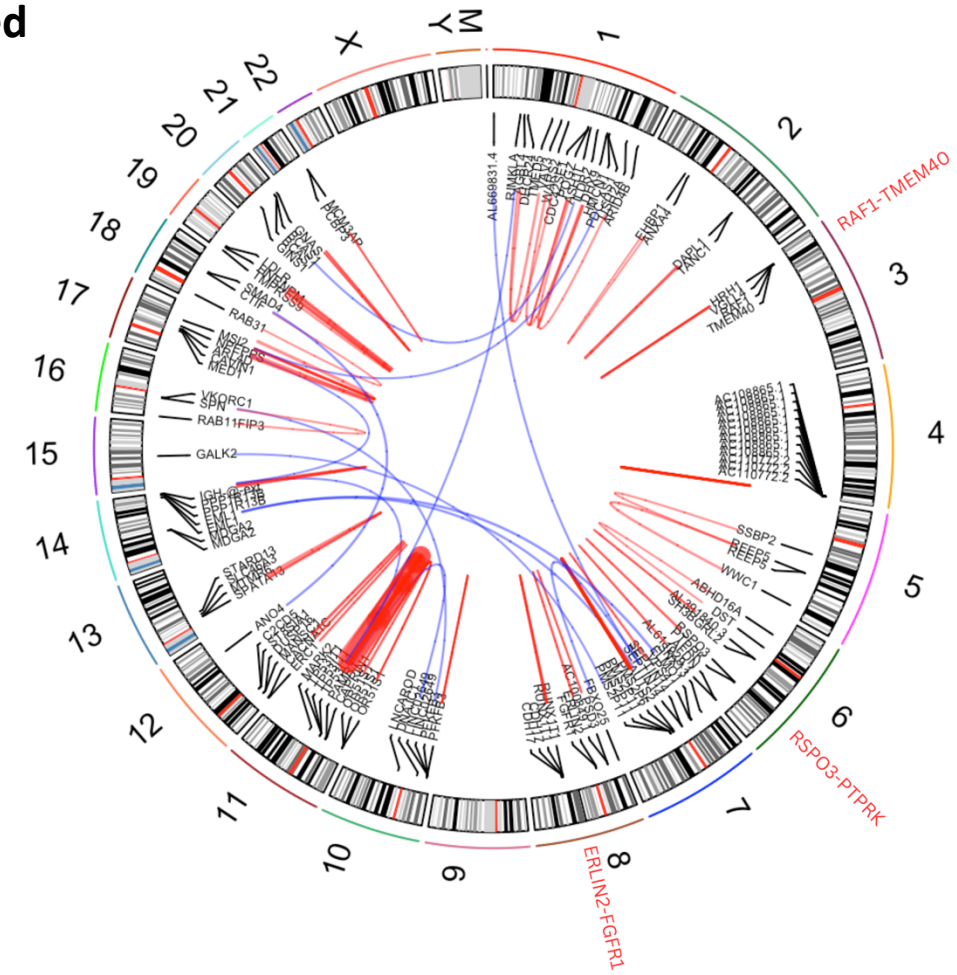


Figure 5: Gene Fusion Analysis stratified by Tumor Mutational Burden (TMB) in CRC Tumors from 67 Hispanic/Latino Patients. This circle plot illustrates the gene fusions detected across all samples from our Hispanic/Latino (H/L) colorectal cancer (CRC) cohort, stratified by hypermutated and non-hypermutated tumors. Clinically actionable gene fusions are highlighted in red, providing insights into potential therapeutic targets for this patient population.

Figure 6: Integrative Genomic Alteration Patterns in Selected Pathways Among CRC Tumors from 67 Hispanic/Latino Patients. This grid represents integrative genomic alteration patterns in the WNT, TGFB, TP53, IGF2/PI3K, RTK/RAS, and combined RTK/RAS/PI3K pathways among colorectal cancer (CRC) tumors from 67 Hispanic/Latino patients. Each column corresponds to an individual case, and each row represents a gene. The grid is divided into hypermutated and non-hypermutated samples, further subdivided by high (>55%) or low (≤55%) 1KG-PEL-like similarity proportion. The color codes represent different attributes: red for amplifications, blue for deletions, a yellow border for early onset (<50 years), a blue border for late onset (≥50 years), and a green dot for somatic mutations.