



# HHS Public Access

Author manuscript

*Cancer*. Author manuscript; available in PMC 2020 June 15.

Published in final edited form as:

*Cancer*. 2019 June 15; 125(12): 2002–2010. doi:10.1002/cncr.31994.

## Clinical and molecular characterization of early-onset colorectal cancer

Alexandra N. Willauer<sup>1</sup>, Yusha Liu, MS<sup>2</sup>, Allan A.L. Pereira, MD, PhD<sup>1</sup>, Michael Lam, MD<sup>1</sup>, Jeffrey S. Morris, PhD<sup>3</sup>, Kanwal P.S. Raghav, MD<sup>1</sup>, Van. K Morris, MD<sup>1</sup>, David Menter, PhD<sup>1</sup>, Russell Broaddus, MD, PhD<sup>4</sup>, Funda Meric-Bernstam, MD<sup>5</sup>, Andrea Hayes-Jordan, MD<sup>6</sup>, Winston Huh, MD<sup>7</sup>, Michael J. Overman, MD<sup>1</sup>, Scott Kopetz, MD, PhD<sup>1</sup>, and Jonathan M. Loree, MD<sup>1</sup>

<sup>1</sup>Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX

<sup>2</sup>Department of Department of Statistics, Rice University, Houston, TX

<sup>3</sup>Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX

<sup>4</sup>Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX

<sup>5</sup>Department of Investigational Cancer Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, TX

<sup>6</sup>Department of Pediatric Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX

<sup>7</sup>Department of Pediatrics, The University of Texas MD Anderson Cancer Center, Houston, TX

### Abstract

---

**Corresponding Author:** Dr. Scott Kopetz, Division of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Unit 426, 1515 Holcombe Blvd, Houston, TX, 77030, Telephone: 713-792-2828, [skopetz@mdanderson.org](mailto:skopetz@mdanderson.org).  
Author Contributions:

**Alexandra Willauer:** conceptualization, data curation, formal analysis, methodology, writing - original draft, and writing - review and editing.

**Yusha Liu:** formal analysis, methodology, writing - review and editing.

**Allan Pereira:** conceptualization, writing - review and editing.

**Michael Lam:** conceptualization, writing - review and editing.

**Jeffrey Morris:** formal analysis, methodology, writing - review and editing.

**Kanwal Raghav:** conceptualization, writing - review and editing.

**Van Morris:** conceptualization, writing - review and editing.

**David Menter:** conceptualization, writing - review and editing.

**Russell Broaddus:** conceptualization, data curation, writing - review and editing.

**Funda Meric-Bernstam:** conceptualization, data curation, writing - review and editing.

**Andrea Hayes-Jordan:** conceptualization, data curation, writing - review and editing.

**Winston Huh:** conceptualization, data curation, writing - review and editing.

**Michael Overman:** conceptualization, writing - review and editing.

**Scott Kopetz:** conceptualization, data curation, formal analysis, methodology, supervision, funding acquisition, writing - original draft, and writing - review and editing.

**Jonathan Loree:** conceptualization, data curation, formal analysis, methodology, supervision, writing - original draft, and writing - review and editing.

**Disclosures:** None

**Purpose**—Colorectal cancer (CRC) incidence is increasing in adults <50 years old. We evaluated clinical and molecular features to identify those unique to early-onset CRC that differentiates these patients from those >50.

**Methods**—Baseline characteristics were evaluated based on CRC onset age using three independent cohorts. A fourth cohort described the impact of age on Consensus Molecular Subtype (CMS) prevalence.

**Results**—Our retrospective review of >36,000 CRC patients showed that early-onset patients were more likely to have microsatellite instability ( $P=0.038$ ), synchronous metastatic disease ( $P=0.009$ ), primary tumors in the distal colon or rectum ( $P<0.0001$ ) and fewer BRAF V600 mutations ( $P<0.001$ ) compared to patients aged  $\geq 50$  years. Patients aged 18–29 had fewer APC mutations (OR 0.56, 95% CI 0.35–0.90,  $P=0.015$ ) and increased prevalence of signet ring histology (OR 4.89, 95% CI 3.23–7.39,  $P<0.0001$ ) compared to other patients <50 years. In patients <40 years, CMS1 was the most common subtype, while CMS3 and CMS4 were uncommon ( $P=0.003$ ). CMS2 was relatively stable across age groups. Early-onset patients with inflammatory bowel disease were more likely to have mucinous or signet ring histology (OR 5.54, 95% CI 2.24–13.74,  $P=0.0004$ ) and less likely to have APC mutations (OR 0.24, 95% CI 0.07–0.75,  $P=0.019$ ) compared to early-onset patients without predisposing conditions.

**Conclusion**—Early-onset CRC is not only distinct from traditional CRC, but special consideration and further investigation should be given to both very young CRC patients (18–29 years) and those with predisposing conditions. The etiology of the high rate of CMS1 in patients <40 deserves further exploration.

## Condensed Abstract

In this manuscript, we demonstrate that early-onset CRC (<50 years of age) has distinct clinical and molecular features including increased prevalence of synchronous metastatic disease, microsatellite instability, primary tumors located in the distal colon or rectum, and CMS1 subtype and fewer mutations in BRAF V600 compared to patients  $\geq 50$  years of age. Among early-onset CRC patients, those aged 18–29 years and those with predisposing conditions were distinct from the other age groups <50, suggesting that considering all CRC patients <50 together may not be appropriate.

## Keywords

colorectal cancer; early-onset; age; Consensus Molecular Subtypes; mutations; CIMP; hereditary; inflammatory bowel disease

## Introduction

Colorectal cancer (CRC) ranks among the top three most prevalent and lethal cancers for men and women.<sup>1</sup> Incidence and mortality for individuals aged  $\geq 50$  years has decreased due to preventive screening, however incidence has increased 1–3% annually for individuals <50 during the same time.<sup>2,3</sup> Though an estimated 4–21% of early-onset CRCs are related to hereditary nonpolyposis colorectal cancer (HNPCC), familial adenomatous polyposis (FAP), or other genetic syndromes, the majority are sporadic.<sup>3–9</sup> Given the discrepant trends in

incidence between early (<50) and late-onset (≥50) CRC and the large number of early-onset patients without identifiable predisposing genetic conditions, further characterization of early-onset CRC is required.

Early-onset CRCs are more likely to exhibit later stage of presentation, distal primary tumors, signet ring histology, and presentation with or development of metastatic disease.<sup>3–7,10,11</sup> From a mutational perspective, early-onset CRC has a lower prevalence of BRAF V600 and NRAS mutations (range 0–8% and 7%, respectively). Most reports describe a lower prevalence of KRAS mutations (range 4–27%) compared to late-onset CRC, however Watson et. al. report a higher prevalence (54%).<sup>3–8,10</sup> Though molecular characterization of these limited gene sets in small populations is available, a comprehensive molecular characterization of early-onset CRC in a large cohort is lacking and may help improve outcomes and an understanding of the unique biology among these patients.

In order to provide a clearer landscape of the clinical and molecular features of early-onset CRC, we performed a retrospective review of >36,000 patients across 4 distinct cohorts that included a mixture of baseline characteristic annotation. Given recent guideline changes by the American Cancer Society that now recommend starting average-risk CRC screening at age 45, we also wanted to determine whether different biologic and clinical characteristics may be present across the spectrum of early onset CRC.<sup>12</sup> Patients were divided into the following age groups and characteristics were compared: 18–29 years, 30–39 years, 40–49 years, 50–59 years, 60–69 years, and ≥70 years.

## Methods

### Patient cohorts

This study utilized 4 separate cohorts described in Figure 1. We retrospectively reviewed 1,877 consecutive metastatic CRC (mCRC) patients with a 46-gene next generation sequencing panel (NGS) at MD Anderson Cancer Center (MDACC). In a second cohort, clinical findings were confirmed in over 32,000 patients from 37 years of MDACC tumor registry data. Patients were cross referenced to prevent duplication. Molecular features were validated in a third cohort of 1,868 CRC patients from the American Association for Cancer Research (AACR) Project Genomics Evidence Neoplasia Information Exchange (GENIE).<sup>13</sup> Since MDACC is a contributing party of AACR GENIE, MDACC data was excluded from GENIE. A fourth cohort including expression data from The Cancer Genome Atlas (TCGA) and from MDACC was utilized to describe the impact of age on Consensus Molecular Subtypes (CMS), which had 626 patients. TCGA CMS classification used the expression platforms that were available in the public data set.<sup>14</sup> Patients from MDACC had CMS classification performed on Agilent microarray data, as previously presented.<sup>15</sup> CMS classification used a previously published and publicly available classifier.<sup>15</sup> Early-onset CRC was defined as CRC diagnosed under the age of 50.

### Molecular annotation

Tumor sequencing of the MDACC molecular cohort, AACR Project GENIE, and TCGA cohorts have been previously described.<sup>14,16–18</sup> For comparison between cohorts, only genes

sequenced on all panels were included in the analysis. For the MDACC molecular cohort, microsatellite instability (MSI) status was retrospectively obtained from medical records for patients who had testing as part of clinical care. MSI status was determined by immunohistochemistry for mismatch repair (MMR) protein deficiency or PCR based assessment of microsatellite status. If either method of detection was abnormal, the tumor was classified as MSI-High (MSI-H).

### Statistical methods

Patients were divided into groups based on age at diagnosis. Right sided tumors occurred between the cecum and up to but not including the splenic flexure. Left sided tumors occurred from the splenic flexure to the rectum. Rectal tumors were considered independently. Categorical variables were compared using the Chi-square or Fisher exact tests as appropriate. Statistical significance was defined as  $P < 0.05$ . A combined P-value for variables in multiple cohorts was generated using Fisher's method for overall significance based on data from the independent cohorts.<sup>19</sup> Odds ratios (OR) and the respective 95% confidence intervals (CI) were calculated for significant variables in very young patients (aged 18–29) compared to other early onset patients (aged 30–49) and for early onset patients with predisposing conditions (inflammatory bowel disease (IBD) or a hereditary syndrome) compared to those without. Data was pooled from all cohorts for analysis of very young patients compared to other early onset patients. Twenty-five of the 46 genes from the NGS panel were not analyzed because  $< 10$  patients in the MDACC molecular cohort had mutations in these genes. A generalized additive model was used to generate prevalence estimates by age (Figures 2 and 3) in order to provide a visual representation of the magnitude of differences across age ranges.

Statistical analyses were performed using Graph Pad Prism version 6.0 (La Jolla, California), SPSS version 24.0 (Armonk, New York), and R studio version 3.4.1 (Boston, Massachusetts).

## Results

### Baseline characteristics

Baseline characteristics are summarized in Table 1 and Supplemental Table 1. More primary tumors were left sided or rectal in patients  $< 50$  years of age (combined  $P < 0.0001$ ). Very young CRC patients (aged 18–29) had a higher prevalence of signet ring histology compared to all age groups (combined  $P = 0.0003$ ) and to other early-onset patients aged 30–49 (OR 4.89, 95% CI 3.23–7.39,  $P < 0.0001$ ). In the MDACC molecular cohort, patients  $< 50$  years were more likely to have synchronous metastatic disease ( $P = 0.009$ ) and be MSI-H ( $P = 0.038$ ), and much of this difference was driven by patients  $< 40$  (these variables not available in other cohorts). Although the majority of patients were Caucasian in all cohorts, a greater proportion of patients  $< 30$  years were Hispanic (combined  $P = 0.0015$ ).

## Mutational landscape (MDACC Molecular and AACR GENIE Cohorts)

Table 2 summarizes the mutation profile of the MDACC molecular and AACR cohorts, and Figure 2 demonstrates mutation prevalence and 95% CIs in the MDACC molecular cohort across the age spectrum.

Fewer APC mutations were observed in patients 18–29 years old compared to other early-onset patients (OR 0.56, 95% CI 0.35–0.90,  $P=0.015$ ), and numerically fewer APC mutations were observed in these very young patients in comparison across all age groups (combined  $P=0.096$ ). CTNNB1 mutations were more common in early-onset CRC patients (combined  $P=0.008$ ); these mutations peaked in 30–39 years (4%) compared to other age groups (range 0–2%,  $P=0.02$ ) in the MDACC cohort and were numerically higher in patients <30 years (13%) compared to other groups (3–7%,  $P=0.054$ ) in the AACR cohort. In the AACR cohort, ATM mutations were highest in patients <30 years old (19%) compared to those older than 30 (range 4–10%,  $P=0.004$ ). The MDACC cohort, utilizing only partial sequencing of the ATM gene, had lower ATM mutation prevalence (2% total) and no association with age. Other key genes in CRC biology, TP53, FBXW7, and SMAD4, were not different by age.

BRAF V600 mutations increased with age from 4% in patients <30 years to a maximum of 14% in patients 70 (combined  $P<0.001$ ). While not statistically significant, KRAS mutation prevalence was numerically lower in patients <30 years (range 37–39%) compared to patients 50 years (range 46–51%, combined  $P=0.41$ ). NRAS mutations did not differ between age groups. Because KRAS, NRAS, and BRAF mutations are mutually exclusive and members of the mitogen-activated protein kinases pathway (MAPK), we evaluated the prevalence of any mutation in this pathway. MAPK pathway mutations differed between early and late-onset CRC (range 56–57% in patients <50 vs. range 60–63% in patients 50, combined  $P=0.020$ ). In both cohorts, these mutations were lowest in patients aged 18–29 (48%) and highest in patients 70 years when evaluated across all age groups (range 65–70%, combined  $P=0.008$ ).

### CMS Subtypes

Figure 3 and Supplemental Table 3 shows CMS subtype prevalence by age in 626 stage I-IV patients with expression data. Subtype significantly differed by age group ( $P=0.0003$ ), with CMS1 being present in the highest proportion for patients <40 (46%) and 70 years (23%) compared to patients aged 40–70 years. CMS2 was relatively stable across age groups. CMS3 and CMS4 were uncommon in patients <40 years (4% and 13%, respectively). Patients <30 years of age were excluded due to limited sample size.

### Patients with predisposing conditions

As shown in Supplemental Table 2, only 28/634 (4%) early-onset mCRC patients from the MDACC molecular cohort had a predisposing condition. Eight patients (1%) had a hereditary syndrome; six patients had HNPCC (Lynch Syndrome), and 2 patients had FAP. The remaining 20 patients (3%) had IBD.

Compared to early-onset mCRC patients without predisposing conditions, those with IBD were more likely to have mucinous or signet ring histologic features (OR 5.54, 95% CI 2.24–13.74 P=0.0004) and had fewer APC mutations (OR 0.24, 95% CI 0.07–0.75, P=0.019). Although not statistically significant, they had numerically fewer left-sided (25% vs 52%, P=0.052) tumors, numerically more TP53 mutations (85% vs 65%, P=0.091), and numerically fewer PIK3CA mutations (0% vs 15%, P=0.057) than other early-onset patients.

Patients with recognized hereditary syndromes were more likely to be MSI-H (OR 27.19, 95% CI 6.65–104.50, P<0.0001) compared to early-onset mCRC patients without predisposing conditions, which is a diagnostic feature of HNPCC. These patients also had an increased prevalence of PIK3CA mutations (OR 5.52, 95% CI 1.58–19.15, P=0.025) and decreased TP53 mutation prevalence (OR 0.00, 95% CI 0.00–0.24, P=0.0002). Signet ring or mucinous histology (38% vs 15%, P=0.11) and right-sided tumors (50% vs 24%, P=0.13) were also numerically higher but not significant in these patients.

## Discussion

This study identified unique clinical and molecular features of early-onset CRC from >36,000 patients and demonstrates a novel continuum of changes that suggests not all patients <50 years old are homogeneous, but that there are additional unique molecular and clinicopathologic features in patients <30 years of age and in patients with predisposing conditions. Early-onset patients were more likely to have synchronous metastatic disease, MSI, and distal primary tumors and less likely to have BRAF V600 mutations compared to patients aged 50. Similarly, patients <40 were predominantly CMS1 or CMS2. We noted that very young patients (<30) were less likely to have mutations in APC and more likely to have signet ring histology. Early-onset CRC also appeared to affect a greater number of Hispanic patients <40. Taken together, early-onset CRC appears to have distinct characteristics and further evaluation of the subgroups among these patients is important.

An important strength of our study was the ability to compare characteristics of early-onset CRC stratified by predisposing conditions. Only a very small number of early-onset patients with metastatic disease (28/634 or 4%) had a recognized hereditary syndrome or IBD in the MDACC molecular cohort. Prior reports suggest that a minority of early-onset CRC is attributed to hereditary syndromes or sporadic DNA MMR deficiency, a feature supported by our study.<sup>3–6,8,9,11,23</sup> Screening of high-risk patients with predisposing hereditary syndromes increases the likelihood of diagnosing cancer at earlier stages. Thus, our study may have identified fewer of these patients because we evaluated patients with metastatic disease in the MDACC molecular cohort. Among patients with early-onset mCRC with IBD, key findings included lower APC and PIK3CA mutation prevalence, yet more frequent mutations in TP53. In IBD-associated cancers, prior literature has described that TP53 mutations occur early, while APC mutations occur later during development.<sup>22</sup> Our findings among patients with predisposing conditions highlights potentially unique biology among these groups but also highlights that currently identified conditions are not present in the vast majority of early-onset patients with metastatic disease.

From a molecular standpoint, we identified unique signaling aberrations that highlight important differences between early and late onset CRC. Although MAPK pathway mutations are common in CRC, early-onset patients showed important differences within this pathway. In patients aged 18–29, BRAF V600 mutations were significantly lower and KRAS mutations were numerically but not significantly lower. Combined MAPK pathway mutations were the lowest in patients aged 18–29 compared to other age groups. Interestingly, we also noted important differences in the WNT pathway. CTNNB1 mutations were more prevalent in early-onset CRC, while APC mutations were numerically lowest among patients <30 and bordered on significance. Increased prevalence of CTNNB1 mutations in younger patients has been described in other cancers, including endometrioid-type endometrial cancer<sup>20</sup> and pediatric medulloblastoma.<sup>21</sup> These findings suggest that early-onset CRC may have alternative aberrations that drive pathogenesis, a feature further highlighted by the differences in Consensus Molecular Subtypes we noted, where patients <40 had a predilection for CMS1/CMS2. Across all ages of CRC, Guinney et al. reported CMS subtype prevalence of 14% CMS1, 37% CMS2, 13% CMS3, and 23% CMS4.<sup>15</sup> In contrast to their combined cohort of 18 CRC data sets, our cohort of merged MDACC and TCGA stage I-IV CRC patients suggested that patients <40 had higher prevalence of CMS1. The CMS1 (or MSI Immune) subtype is defined in part by MSI, CIMP, hypermutation, and immune infiltration and activation. HNPCC is defined by high MSI and DNA MMR deficiency, and IBD-associated CRC stems from long duration of colonic inflammation.<sup>22</sup> Thus, predisposing features expected in younger patients may impact the expression patterns within tumors and could contribute to the predilection for CMS1. Alternatively, the high rate of CMS1 subtype may be driven by some yet unknown set of factors in early-onset CRC.

While our study represents one of the largest cohorts of early-onset CRC, it must be interpreted in the context of its limitations. Besides the MDACC molecularly annotated cohort, a number of clinical or molecular variables were missing from each of the 3 validation cohorts that limited our ability to cross compare results. For example, information about predisposing conditions was only available for the MDACC molecular cohort, and it was not feasible to obtain this data for the other cohorts. Additionally, there were only 28 patients with predisposing conditions in our analysis. Our analysis was also hindered by potential false discovery from the investigation of a large number of variables. A Bonferroni correction adjusted analysis for mutations would require  $P < 0.002$  for statistical significance as 21 genes were tested. In this conservative analysis, only mutations in BRAF V600 would be considered significantly different by age. Despite these limitations, this study utilized multiple, large cohorts and evaluated a comprehensive panel of clinical and molecular features, allowing one of the most complete analyses of tumor biology in early-onset CRC to date.

## Conclusion

Early-onset CRC has distinct clinical and molecular features, and it may be more appropriate to consider age as a continuum when evaluating CRC. Early-onset CRC in patients aged 18–29 years was shown to be unique and patients with IBD may also have tumors with unique biology. These notable differences in very young CRC patients and

those with predisposing conditions highlight that early-onset CRC has unique subsets within the population of patients < 50 years of age.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

**Funding:** AW was the recipient of a summer fellowship supported by NCI grants R25E CA056452 and P30 CA016672. SK is the recipient of NIH R01 grants that supported this research (CA172670 & CA187238). This work was also supported by the MD Anderson Colorectal Cancer Moon Shot Program.

## References

1. Edwards BK, Noone AM, Mariotto AB, et al. Annual Report to the Nation on the status of cancer, 1975–2010, featuring prevalence of comorbidity and impact on survival among persons with lung, colorectal, breast, or prostate cancer. *Cancer*. 2014;120(9):1290–1314. doi:10.1002/cncr.28509. [PubMed: 24343171]
2. Siegel RL, Fedewa SA, Anderson WF, et al. Colorectal Cancer Incidence Patterns in the United States, 1974–2013. *JNCI J Natl Cancer Inst*. 2017;109(8):27–32. doi:10.1093/jnci/djw322.
3. Ahnen DJ, Wade SW, Jones WF, et al. The increasing incidence of young-onset colorectal cancer: A call to action. *Mayo Clin Proc*. 2014;89(2):216–224. doi:10.1016/j.mayocp.2013.09.006. [PubMed: 24393412]
4. Watson R, Liu T-C, Ruzinova MB. High frequency of KRAS mutation in early onset colorectal adenocarcinoma: implications for pathogenesis. *Hum Pathol*. 2016;56:163–170. doi:10.1016/j.humpath.2016.06.010. [PubMed: 27346571]
5. Goel A, Nagasaka T, Spiegel J, Meyer R, Lichliter WE, Boland CR. Low frequency of Lynch syndrome among young patients with non-familial colorectal cancer. *Clin Gastroenterol Hepatol*. 2010;8(11):966–971. doi:10.1016/j.cgh.2010.06.030. [PubMed: 20655395]
6. Yantiss RK, Goodarzi M, Zhou XK, et al. Clinical, pathologic, and molecular features of early-onset colorectal carcinoma. *Am J Surg Pathol*. 2009;33(4):572–582. doi:10.1097/PAS.0b013e31818afd6b. [PubMed: 19047896]
7. Chang DT, Pai RK, Rybicki LA, et al. Clinicopathologic and molecular features of sporadic early-onset colorectal adenocarcinoma: an adenocarcinoma with frequent signet ring cell differentiation, rectal and sigmoid involvement, and adverse morphologic features. *Mod Pathol*. 2012;25(8):1128–1139. doi:10.1038/modpathol.2012.61. [PubMed: 22481281]
8. Alsop K, Mead L, Smith LD, et al. Low somatic K-ras mutation frequency in colorectal cancer diagnosed under the age of 45 years. *Eur J Cancer*. 2006;42(10):1357–1361. doi:10.1016/j.ejca.2006.02.023. [PubMed: 16765042]
9. Pearlman R, Frankel WL, Swanson B, et al. Prevalence and Spectrum of Germline Cancer Susceptibility Gene Mutations Among Patients With Early-Onset Colorectal Cancer. *JAMA Oncol*. 2017;3(4):464. doi:10.1001/jamaoncol.2016.5194. [PubMed: 27978560]
10. Kirzin S, Marisa L, Guimbaud R, et al. Sporadic early-onset colorectal cancer is a specific sub-type of cancer: A morphological, molecular and genetics study. *PLoS One*. 2014;9(8). doi:10.1371/journal.pone.0103159.
11. Chen FW, Sundaram V, Chew TA, Ladabaum U. Advanced-Stage Colorectal Cancer in Persons Younger Than 50 Years Not Associated With Longer Duration of Symptoms or Time to Diagnosis. *Clin Gastroenterol Hepatol*. 2017;15(5):728–737.e3. doi:10.1016/j.cgh.2016.10.038. [PubMed: 27856366]
12. Wolf AMD, Fontham ETH, Church TR, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. *CA Cancer J Clin*. 5 2018. doi:10.3322/caac.21457.



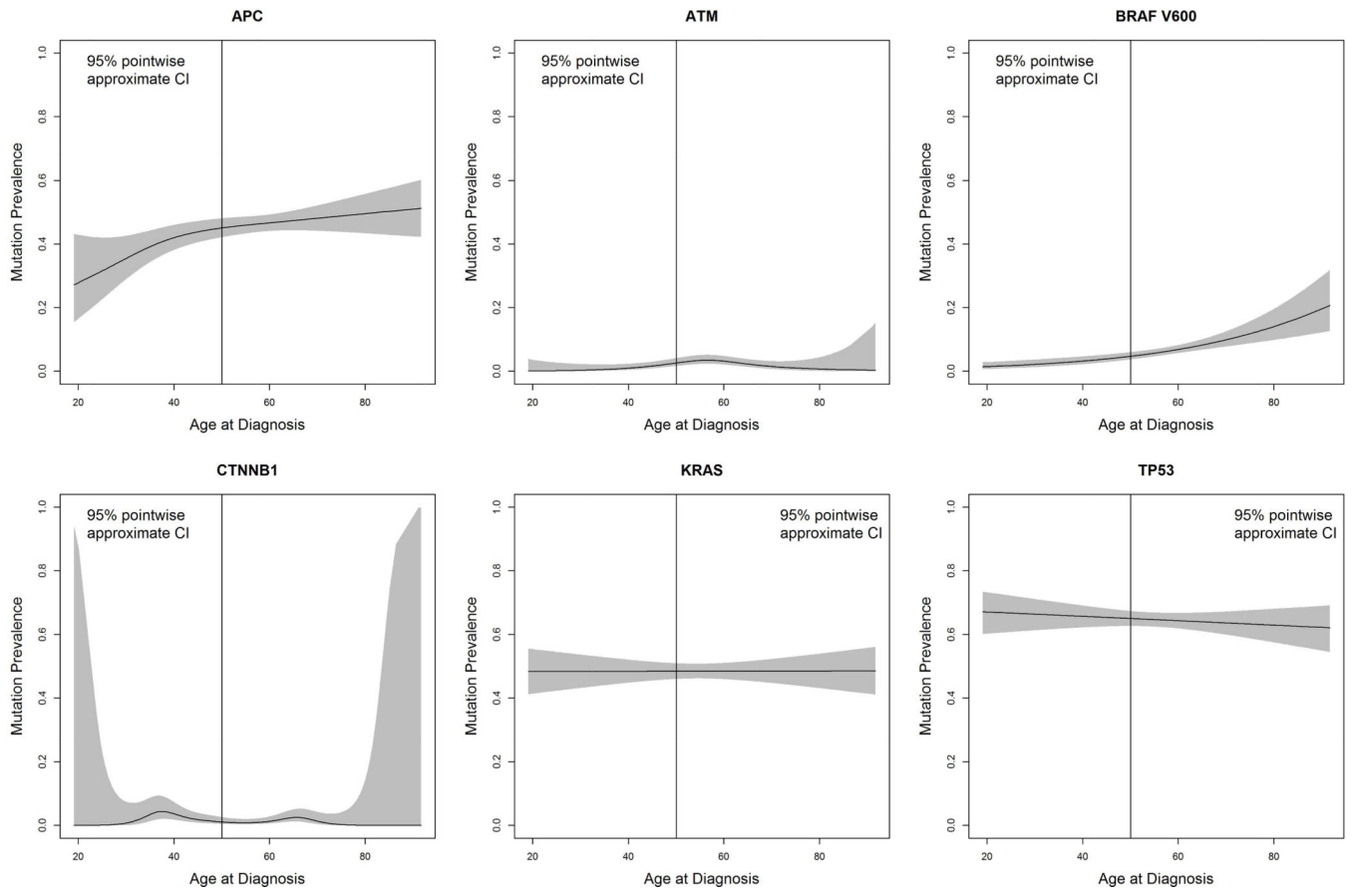
13. The AACR Project GENIE Consortium. AACR Project GENIE: Powering Precision Medicine Through An International Consortium. *Cancer Discov.* 2017. doi:10.1158/2159-8290.CD-17-0151.
14. Cancer Genome Atlas Network, Muzny DM, Bainbridge MN, et al. Comprehensive molecular characterization of human colon and rectal cancer. *Nature.* 2012;487(7407):330–337. doi:10.1038/nature11252. [PubMed: 22810696]
15. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med.* 2015;21(11):1350–1356. doi:10.1038/nm.3967. [PubMed: 26457759]
16. Loree JM, Pereira AAL, Lam M, et al. Classifying colorectal cancer by tumor location rather than sidedness highlights a continuum in mutation profiles and consensus molecular subtypes. *Clin Cancer Res.* 2018;24(5):1062–1072. doi:10.1158/1078-0432.CCR-17-2484. [PubMed: 29180604]
17. Singh RR, Patel KP, Routbort MJ, et al. Clinical Validation of a Next-Generation Sequencing Screen for Mutational Hotspots in 46 Cancer-Related Genes. *J Mol Diagnostics.* 2013;15(5):607–622. doi:10.1016/j.jmoldx.2013.05.003.
18. American Association for Cancer Research. AACR GENIE Data Guide. <http://www.aacr.org/Documents/GENIEDataGuide.pdf>.
19. Fischer RA. *Statistical Methods for Research Workers.* Edinburg: Oliver and Boyd; 1925 <http://www.yorku.ca/pclassic/Fisher/Methods/>.
20. Kurnit KC, Kim GN, Fellman BM, et al. CTNNB1 (beta-catenin) mutation identifies low grade, early stage endometrial cancer patients at increased risk of recurrence. *Mod Pathol.* 2017;30(7):1032–1041. doi:10.1038/modpathol.2017.15. [PubMed: 28281553]
21. Taylor MD, Northcott PA, Korshunov A, et al. Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol.* 2012;123(4):465–472. doi:10.1007/s00401-011-0922-z. [PubMed: 22134537]
22. Ullman TA, Itzkowitz SH. Intestinal Inflammation and Cancer. *Gastroenterology.* 2011;140(6):1807–1816.e1. doi:10.1053/j.gastro.2011.01.057. [PubMed: 21530747]
23. Chang DT, Pai RK, Rybicki LA, et al. Clinicopathologic and molecular features of sporadic early-onset colorectal adenocarcinoma: an adenocarcinoma with frequent signet ring cell differentiation, rectal and sigmoid involvement, and adverse morphologic features. *Mod Pathol.* 2012;25(8):1128–1139. doi:10.1038/modpathol.2012.61. [PubMed: 22481281]

	MDACC Molecular Cohort	MDACC Tumor Registry Cohort	AACR Project GENIE Cohort	CMS Cohort
Patient Information	<ul style="list-style-type: none"> <li>N=1877</li> <li>Seen at MDACC from January 1, 2012 to September 1, 2016</li> </ul>	<ul style="list-style-type: none"> <li>N=32507</li> <li>Seen at MDACC from January 1, 1980 to present</li> </ul>	<ul style="list-style-type: none"> <li>N=1868</li> <li>Excluded patients from MDACC to prevent duplication of data</li> </ul>	<ul style="list-style-type: none"> <li>Total N=626</li> <li>N=448 from TCGA</li> <li>N=178 from MDACC</li> </ul>
Clinical Data	<ul style="list-style-type: none"> <li>Baseline clinical and pathologic characteristics</li> </ul>	<ul style="list-style-type: none"> <li>Baseline clinical and pathologic characteristics</li> </ul>	<ul style="list-style-type: none"> <li>Limited clinical and pathologic characteristics</li> </ul>	<ul style="list-style-type: none"> <li>Limited clinical and pathologic characteristics</li> </ul>
Molecular Data	<ul style="list-style-type: none"> <li>Mutational data available from 46 or 50 gene CLIA next generation sequencing panel</li> </ul>	<ul style="list-style-type: none"> <li>Unavailable</li> </ul>	<ul style="list-style-type: none"> <li>Mutation data available from AACR Project GENIE database includes a mixture of next generation sequencing platforms</li> </ul>	<ul style="list-style-type: none"> <li>RNA expression data.</li> <li>For TCGA patients, data was publicly available.</li> <li>For MDACC patients, data was obtained using Affymetrix RNA expression arrays.</li> </ul>
Cancer Stage(s)	<ul style="list-style-type: none"> <li>Stage IV</li> </ul>	<ul style="list-style-type: none"> <li>Stages I-IV</li> </ul>	<ul style="list-style-type: none"> <li>Majority stage IV</li> </ul>	<ul style="list-style-type: none"> <li>Stages I-IV</li> </ul>
Additional Data	<ul style="list-style-type: none"> <li>Co-morbid predisposing condition information available for patients &lt; 50 years</li> </ul>			<ul style="list-style-type: none"> <li>Classification by CMS subtype</li> </ul>

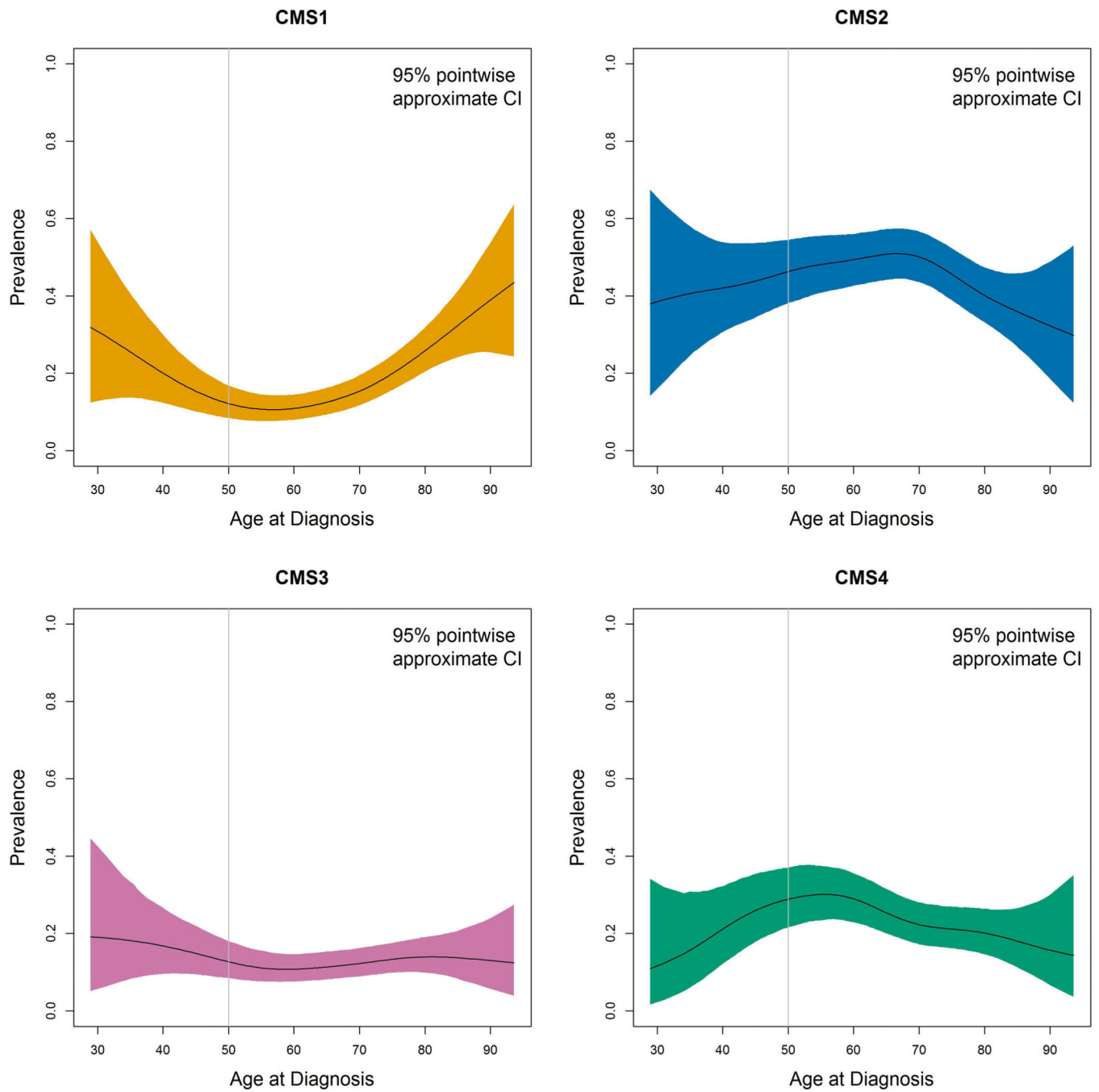
**Figure 1:**

Cohorts utilized in this study.

Sub-caption: MDACC = MD Anderson Cancer Center. AACR = American Association for Cancer Research. GENIE = Genomics Evidence Neoplasia Information Exchange. CMS = Consensus Molecular Subtypes. TCGA = The Cancer Genome Atlas. CRC = colorectal cancer.



**Figure 2:** Mutation prevalence in metastatic colorectal cancer patients according to age at diagnosis in MDACC molecular cohort. Sub-caption: CI = confidence interval. Total number (N) = 1,877.



**Figure 3:**  
Prevalence of Consensus Molecular Subtypes (CMS) by age.  
Sub-caption: CI = confidence interval. Total number (N) = 626; N for CMS1=110,  
CMS2=287, CMS3=79, CMS4=149.

**Table 1:**

Baseline characteristics of the MDACC molecular cohort classified by age

Age (years)	18–29	30–39	40–49	50–59	60–69	70	P
<b>N (%)</b>	46 (2)	177 (9)	411 (22)	605 (32)	454 (24)	184 (10)	
<b>Gender</b>							
<b>Male</b>	20 (44)	94 (53)	208 (51)	35 (58)	263 (58)	123 (67)	<b>0.002</b>
<b>Female</b>	26 (57)	83 (47)	203 (49)	255 (42)	191 (42)	61 (33)	
<b>Race/Ethnicity</b>							
<b>Caucasian</b>	33 (72)	121 (68)	299 (73)	456 (75)	332 (73)	146 (80)	0.094*
<b>African American</b>	2 (4)	20 (11)	31 (8)	68 (11)	45 (10)	16 (9)	
<b>Hispanic</b>	9 (20)	22 (12)	49 (12)	54 (9)	49 (11)	16 (9)	
<b>Other</b>	2 (4)	14 (8)	32 (8)	27 (5)	28 (6)	6 (3)	
<b>Location (N=1876, 99% known)</b>							
<b>Right Colon</b>	8 (17)	41 (23)	109 (27)	192 (33)	183 (40)	71 (39)	< <b>0.001</b>
<b>Left Colon</b>	24 (52)	94 (53)	209 (51)	275 (46)	179 (39)	69 (38)	
<b>Rectum</b>	14 (30)	42 (24)	93 (23)	138 (23)	92 (20)	43 (23)	
<b>Metastatic Disease</b>							
<b>Metachronous</b>	14 (30)	49 (28)	133 (32)	206 (34)	183 (40)	77 (42)	<b>0.009</b>
<b>Synchronous</b>	32 (70)	128 (72)	278 (68)	399 (66)	271 (60)	107 (58)	
<b>Histology (N=1871, 99% known)</b>							
<b>Non-mucinous, Non-signet ring Adenocarcinoma</b>	35 (80)	146 (83)	344 (84)	506 (84)	367 (81)	151 (81)	0.60
<b>Mucinous</b>	6 (13)	25 (14)	61 (15)	84 (14)	72 (16)	29 (16)	
<b>Signet ring</b>	3 (7)	5 (3)	5 (1)	14 (2)	14 (3)	4 (2)	
<b>MSI-H (N=1525, 81% known)</b>	3 (7)	12 (8)	23 (3)	11 (2)	13 (4)	6 (4)	<b>0.038</b>

Values represent number followed in brackets by % of known patients. N = number.

\* Analysis of race/ethnicity excluding “other” yielded P=0.19

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2:** Mutation profile in the MDACC molecular and AACR Project GENIE cohorts classified by age

Age (years)	MDACC Molecular Cohort							AACR Project GENIE Cohort							P	Combined P
	18-29	30-39	40-49	50-59	60-69	70	P	18-29	30-39	40-49	50-59	60-69	70	P		
N (%)	46 (2)	177 (9)	411 (22)	605 (32)	454 (24)	184 (10)		31 (2)	126 (7)	371 (20)	518 (28)	510 (27)	312 (17)			
Mutation																
APC	12 (26)	71 (40)	181 (44)	287 (47)	208 (46)	88 (48)	0.059	18 (58)	81 (64)	245 (66)	338 (65)	303 (59)	196 (63)	0.33	0.096	
AKT1	1 (2)	3 (2)	1 (0)	4 (1)	4 (1)	1 (1)	0.25	1 (3)	2 (2)	6 (2)	2 (0)	9 (2)	8 (3)	0.067	0.085	
ATM	0 (0)	1 (1)	7 (2)	23 (4)	9 (2)	1 (1)	<b>0.037</b>	6 (19)	5 (4)	25 (7)	31 (6)	23 (5)	31 (10)	<b>0.004</b>	<b>0.001</b>	
BRAF	2 (4)	8 (5)	27 (7)	36 (6)	50 (11)	21 (11)	<b>0.006</b>	4 (13)	16 (13)	33 (9)	46 (9)	59 (12)	55 (18)	<b>0.004</b>	< <b>0.001</b>	
BRAF V600	2 (4)	5 (3)	17 (4)	30 (5)	43 (10)	18 (10)	<b>0.001</b>	0 (0)	11 (9)	21 (6)	31 (6)	45 (9)	42 (14)	<b>0.001</b>	< <b>0.001</b>	
CDKN2A	1 (2)	2 (1)	1 (0)	2 (0)	3 (1)	3 (2)	0.11	0 (0)	3 (2)	10 (3)	7 (1)	11 (2)	10 (3)	0.51	0.22	
CTNNB1	0 (0)	7 (4)	8 (2)	5 (1)	10 (2)	0 (0)	<b>0.020</b>	4 (13)	9 (7)	10 (3)	23 (4)	23 (5)	18 (6)	0.054	<b>0.008</b>	
ERBB2	0 (0)	2 (1)	3 (1)	7 (1)	4 (1)	4 (2)	0.71	3 (10)	5 (4)	12 (3)	16 (3)	17 (3)	11 (4)	0.53	0.74	
ERBB4	0 (0)	1 (1)	4 (1)	12 (2)	4 (1)	1 (1)	0.55	3 (10)	7 (6)	20 (5)	18 (4)	27 (5)	14 (5)	0.40	0.55	
FGFR3	1 (2)	2 (1)	3 (1)	3 (1)	2 (0)	0 (0)	0.37	2 (7)	3 (2)	11 (3)	9 (2)	10 (2)	13 (4)	0.15	0.22	
FBXW7	2 (4)	8 (5)	34 (8)	50 (8)	29 (6)	19 (10)	0.26	2 (7)	10 (8)	33 (9)	57 (11)	46 (9)	41 (13)	0.36	0.31	
GNAS	0 (0)	1 (1)	5 (1)	10 (2)	9 (2)	5 (3)	0.61	2 (7)	6 (5)	17 (5)	8 (2)	23 (5)	11 (4)	<b>0.031</b>	0.094	
KDR	0 (0)	1 (1)	5 (1)	6 (1.0)	4 (1)	3 (2)	0.91	3 (10)	3 (2)	13 (4)	10 (2)	18 (4)	10 (3)	0.18	0.46	
KIT	0 (0)	1 (1)	5 (1)	5 (1)	2 (0)	0 (0)	0.68	1 (3)	0 (0)	14 (4)	8 (2)	11 (2)	4 (1)	0.071	0.19	
KRAS	17 (37)	89 (50)	207 (50)	292 (48)	210 (46)	94 (51)	0.46	12 (39)	47 (37)	161 (43)	244 (47)	239 (47)	148 (47)	0.30	0.41	
MET	1 (2)	1 (1)	1 (0)	5 (1)	3 (1)	1 (1)	0.52	1 (3)	5 (4)	12 (3)	7 (1)	4 (1)	8 (3)	<b>0.023</b>	0.065	
NRAS	3 (7)	8 (5)	14 (3)	22 (4)	24 (5)	8 (4)	0.61	1 (3)	5 (4)	15 (4)	20 (4)	26 (5)	17 (6)	0.87	0.87	
PIK3CA	4 (9)	27 (15)	66 (16)	80 (13)	72 (16)	38 (21)	0.16	8 (26)	29 (23)	67 (18)	102 (20)	94 (18)	58 (19)	0.71	0.36	
PTEN	2 (4)	2 (1)	14 (3)	6 (1)	16 (4)	6 (3)	<b>0.017</b>	1 (3)	8 (6)	28 (8)	23 (4)	28 (6)	14 (5)	0.42	0.042	
RBI	1 (2)	1 (1)	6 (2)	2 (0)	4 (1)	2 (1)	0.24	2 (7)	4 (3)	12 (3)	11 (2)	12 (2)	7 (2)	0.53	0.39	
RET	0 (0)	1 (1)	4 (1)	7 (1)	2 (0)	2 (1)	0.85	3 (10)	4 (3)	12 (3)	6 (1)	11 (2)	6 (2)	<b>0.036</b>	0.14	
SMAD4	9 (20)	24 (14)	53 (13)	66 (11)	58 (13)	23 (13)	0.58	5 (16)	14 (11)	51 (14)	63 (12)	72 (14)	41 (13)	0.88	0.85	
SMARCB1	0 (0)	0 (0)	4 (1)	4 (1)	6 (1)	1 (1)	0.71	2 (7)	2 (2)	6 (2)	4 (1)	2 (0)	8 (3)	<b>0.010</b>	0.042	
SMO	0 (0)	0 (0)	4 (1)	5 (1)	6 (1)	1 (1)	0.77	4 (13)	6 (5)	12 (3)	12 (2)	9 (2)	7 (2)	<b>0.020</b>	0.080	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	MDACC Molecular Cohort						AACR Project GENIE Cohort							
	0 (0)	1 (1)	2 (1)	8 (1)	1 (0)	2 (1)	0 (0)	1 (1)	2 (1)	3 (1)	12 (2)	4 (1)	0.14	
<b>STK11</b>	28 (61)	120 (68)	265 (65)	398 (66)	280 (62)	123 (67)	18 (58)	91 (72)	269 (73)	347 (67)	326 (64)	204 (65)	0.06	0.16
<b>TP53</b>	22 (48)	102 (58)	239 (58)	349 (58)	274 (60)	120 (65)	15 (48)	71 (56)	211 (57)	303 (58)	321 (63)	217 (70)	<b>0.004</b>	<b>0.008</b>
<b>MAPK Summary</b>														

Only genes mutated in >10 patients in the MDACC molecular cohort are displayed. Values represent number followed in brackets by %. N = number.