scientific reports

OPEN

Molecular drivers of tumor progression in microsatellite stable *APC* mutation-negative colorectal cancers

Adam Grant[1], Rosa M. Xicola[2], Vivian Nguyen[1], James Lim[3], Curtis Thorne[4], Bodour Salhia[5], Xavier Llor[2], Nathan Ellis[4] & Megha Padi[1]✉

The tumor suppressor gene adenomatous polyposis coli (*APC*) is the initiating mutation in approximately 80% of all colorectal cancers (CRC), underscoring the importance of aberrant regulation of intracellular WNT signaling in CRC development. Recent studies have found that early-onset CRC exhibits an increased proportion of tumors lacking an *APC* mutation. We set out to identify mechanisms underlying *APC* mutation-negative (APC^{mut-}) CRCs. We analyzed data from The Cancer Genome Atlas to compare clinical phenotypes, somatic mutations, copy number variations, gene fusions, RNA expression, and DNA methylation profiles between APC^{mut-} and *APC* mutation-positive (APC^{mut+}) microsatellite stable CRCs. Transcriptionally, APC^{mut-} CRCs clustered into two approximately equal groups. Cluster One was associated with enhanced mitochondrial activation. Cluster Two was strikingly associated with genetic inactivation or decreased RNA expression of the WNT antagonist *RNF43*, increased expression of the WNT agonist *RSPO3*, activating mutation of *BRAF*, or increased methylation and decreased expression of *AXIN2*. APC^{mut-} CRCs exhibited evidence of increased immune cell infiltration, with significant correlation between M2 macrophages and *RSPO3*. APC^{mut-} CRCs comprise two groups of tumors characterized by enhanced mitochondrial activation or increased sensitivity to extracellular WNT, suggesting that they could be respectively susceptible to inhibition of these pathways.

Colorectal cancer (CRC) is the second deadliest cancer in the United States, with an estimated 147,950 individuals diagnosed and 53,200 deaths in 2020[1]. Although there have been great reductions in CRC incidence and mortality widely attributed to increased screening[2], the incidence of CRC has been increasing in individuals less than 50 years of age at a rate of 2% per year since 1994[3]. Molecular analysis has shown that < 20% of early-onset CRC cases are explained by genetically determined hereditary syndromes[4] and a variety of environmental factors have been postulated to underlie its increase[5], suggesting that a unitary cause of early-onset CRC will be elusive. With early-onset CRC manifesting as a heterogenous disease caused by a multitude of factors, there is a pressing need to identify the distinct molecular subtypes of CRC that are overrepresented in early-onset cases.

Somatic mutation of the adenomatous polyposis coli (*APC*) gene is the initiating event in approximately 80% of all CRCs, but *APC* mutations are significantly less frequent in early-onset CRCs[6–8]. *APC* is a structural and regulatory component of a destruction complex, which responds to WNT stimulation by inhibition of degradation of the stem cell-promoting transcription factor β-catenin, encoded by the *CTNNB1* gene[9]. Failure to regulate β-catenin by degradation due to mutational inactivation of *APC* throws colorectal epithelial cells into a continuous "WNT-activated" state; they no longer require activation by WNTs to maintain the stem cell compartment[10]. The fact that early-onset CRCs more frequently lack an *APC* mutation suggests that many of these tumors depend on alternative molecular mechanisms. In mismatch repair-deficient CRCs, which exhibit microsatellite instability (MSI) and constitute 12–15% of all CRCs, *APC* mutations are also significantly less frequent and *BRAF* mutations constitute a dominant driver mechanism[11]. What initiates and drives the carcinogenetic process in microsatellite

[1]University of Arizona Cancer Center, University of Arizona, 1515 N. Campbell Avenue, Tucson, AZ 85724, USA. [2]Department of Medicine and Cancer Center, Yale University, New Haven, CT, USA. [3]Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ, USA. [4]Department of Cellular and Molecular Medicine, University of Arizona, Tucson, AZ, USA. [5]Department of Translational Genomics, University of Southern California, Los Angeles, CA, USA. ✉email: mpadi@arizona.edu

stable (MSS) CRCs that lack *APC* mutations? Here, we comprehensively compare molecular profiles of MSS *APC* mutation-positive CRCs (APC^{mut+}) and MSS *APC* mutation-negative (APC^{mut-}) CRCs to identify novel *APC*-independent mechanisms driving CRC subtypes.

Methods

Analyses of genomic alterations in case series. To formulate a discovery series, we obtained colon adenocarcinoma (COAD) and rectal adenocarcinoma (READ) data from The Cancer Genome Atlas (TCGA) in the Genomic Data Commons portal (Supplementary File 1). Curated somatic nucleotide variant and copy number data were extracted using TCGAbiolinks and FireBrowse, respectively[12]. Deep deletions, amplifications, and gene fusions[13] were identified. We excluded hypermutable cases by removing MSI-high cases based on clinical data and by removing cases with > 700 mutations. CRC samples were classified as APC^{mut-} if they lacked a non-silent mutation or deep (homozygous) deletion in *APC* or lacked a mutation in *CTNNB1*[14]. With these filtration steps, we had 63 APC^{mut-} samples and 362 APC^{mut+} samples. We compared genomic alterations between APC^{mut-} and APC^{mut+} CRCs by Fisher's exact test and tested mutual exclusivity by CoMEt[15]. For more details on the bioinformatics analysis, see Supplementary Methods.

For validation series, we used the CPTAC-2[16] and GSE35896[17] datasets, because they were the only CRC datasets with *APC* mutation status and gene expression data available from the cBioPortal, the International Cancer Genome Consortium or studies utilized by Guinney et al. to determine consensus molecular subtypes of CRC[18]. CPTAC-2 was downloaded from cBioPortal and GSE3896 from synapse.org[18]. In the CPTAC-2 dataset, we identified 11 APC^{mut-} CRCs and 70 APC^{mut+} CRCs. Because GSE35896 did not have whole exome sequencing or copy number data, we could not filter for hypermutation, *APC* deep deletion, or *CTNNB1* mutations. Based on the data available, we classified 16 out of 56 MSS CRC samples as APC^{mut-}.

Transcriptomic analyses. For TCGA and CPTAC-2 series, we obtained HTSeq count files and used the edgeR and limma pipeline to normalize the counts matrix. For GSE35896, we used the RMA normalized data. Genes that had less than one count per million in more than half the samples were discarded. MBatch analysis showed no evidence of batch effects.

We used limma to identify differentially expressed genes between APC^{mut-} and APC^{mut+} CRCs ($P_{adj} < 0.05$). PathView was used to map differentially expressed genes onto the KEGG WNT canonical signaling pathway, with the node sum parameter set to "max.abs"[19]. Gene set enrichment analysis (GSEA) was performed using fgsea[20]. For GSEA input, we used the Gene Ontology (GO) biological process gene sets from MSigDB and ranked the genes by the t-statistic from our differential expression analysis. The Cytoscape application EnrichmentMap was used to visualize all statistically significant GO terms ($P_{adj} < 0.05$)[21]. CIBERSORTx was used to impute the fraction of immune cells based on gene expression data from the TCGA, GSE35896, and CPTAC-2 datasets[22].

To characterize the theoretical WNT ligand sensitivity of APC^{mut-} tumors, we defined a score by the normalized expression of *RSPO3* minus the sum of the normalized expression levels of *RNF43* and *ZNRF3*:

$$\mathrm{WNT_{LS}} = \mathrm{Z}[RSPO3_{mRNA-z} - (RNF43_{mRNA-z} + ZNRF3_{mRNA-z})],$$

where the subscript mRNA-z indicates the z-score of the expression value relative to all samples including tumors and normals. The WNT_{LS} score for each tumor was then compared to the maximum WNT ligand expression over all 12 WNTs in the same tumor.

DNA methylation analyses. Methylation was assayed by TCGA using Illumina Human Methylation 450 arrays and data was accessed using TCGAbiolinks. Preprocessing and normalization were carried out with the R package minfi[23]. MBatch analysis showed no evidence of batch effects. Differentially methylated regions (DMRs) were identified using DMRcate and annotated with annotatr[24,25]. For DMRs that spanned multiple gene regions, we selected the gene with the most significant beta-values. To quantify methylation of a DMR, we took the average of all the statistically significant beta-values associated with the DMR.

DepMap data analyses. DepMap data was obtained from https://depmap.org/portal/download/[26]. CRC cell lines were selected excluding those with MSI and with > 800 mutations. To distinguish APC^{mut-} from APC^{mut+} CRC cell lines, we used the same filtering steps we used for the TCGA dataset. To assess the effect of CRISPR knockouts, we applied a Welch's two-sample t-test statistic to the dependency scores of APC^{mut-} and APC^{mut+} cell lines. Dependency scores were extracted from the file "Achilles_gene_dependency.csv" on the DepMap portal.

Ethics statement. Ethics approval is not required for this study because it does not involve human participants or animal subjects.

Results

Age effect in APC^{mut-} CRCs. To identify characteristics that distinguished APC^{mut-} from APC^{mut+} CRCs, we compared molecular profiles between the two groups in a discovery series from the TCGA, then validated the results in two additional publicly available series. CRC samples that exhibited MSI or were hypermutated were excluded from our study, because tumors with these characteristics constitute a well-defined subtype[11]. In addition to separating MSS and non-hypermutated CRCs by *APC* mutation status, samples that contained a *CTNNB1* mutation[14] or deep deletion of *APC* were also classified as APC^{mut+}. After applying these filtration steps, we classified 63 of 425 (15%) of the MSS CRCs in the TCGA dataset as APC^{mut-}. In the GSE35896 valida-

Feature	APC^{mut+} (N = 362, 85%)	APC^{mut-} (N = 63, 15%)	*P* value
Age	66.4	61.4	**.004**
Non-silent mutations	121.4	112.4	.21
Male/female	194/165 (54%)	30/33 (48%)	.41
COAD/READ	250/109 (70%)	51/12 (81%)	.07
Proximal/distal	104/195 (35%)	24/25 (49%)	.08
White	305/359 (85%)	53/63 (84%)	.85
African American	46/359 (13%)	5/63 (8%)	.40
Asian	4/359 (1%)	5/63 (8%)	**.005**
American Indian	4/359 (1%)	0/63 (0%)	1.0
Stage I	62/346 (18%)	4/59 (7%)	**.035**
Stage II	106/346 (31%)	23/59 (39%)	.23
Stage III	119/346 (34%)	18/59 (31%)	.66
Stage IV	59/346 (17%)	14/59 (23%)	.27
CIMP-0	177/251 (70%)	30/50 (60%)	.18
CIMP-low	57/251 (23%)	11/50 (22%)	1.0
CIMP-high	17/234 (7%)	9/50 (18%)	**.02**
CMS1	5/291 (2%)	2/49 (4%)	.27
CMS2	149/291 (51%)	20/49 (41%)	.22
CMS3	41/290 (14%)	7/49 (14%)	1.0
CMS4	96/290 (33%)	20/49 (41%)	.33

Table 1. Comparison of clinical features in *APC* mutation-positive (APC^{mut+}) and *APC* mutation-negative (APC^{mut-}) colorectal cancers. *P* values were calculated for comparisons between APC^{mut+} CRCs and APC^{mut-} CRCs from the TCGA dataset. A two-sample t-test with a two-tailed *p* value was performed for continuous features and a Fisher's exact test with a two-tailed *p* value was performed for categorical data. A *p* value threshold of 0.05 was considered significant. CIMP, CpG island methylator phenotype was defined by unsupervised clustering as reported byGuinney et al. CMS, consensus molecular subtypes of colorectal cancer determined by Guinney et al. Significant values are in bold.

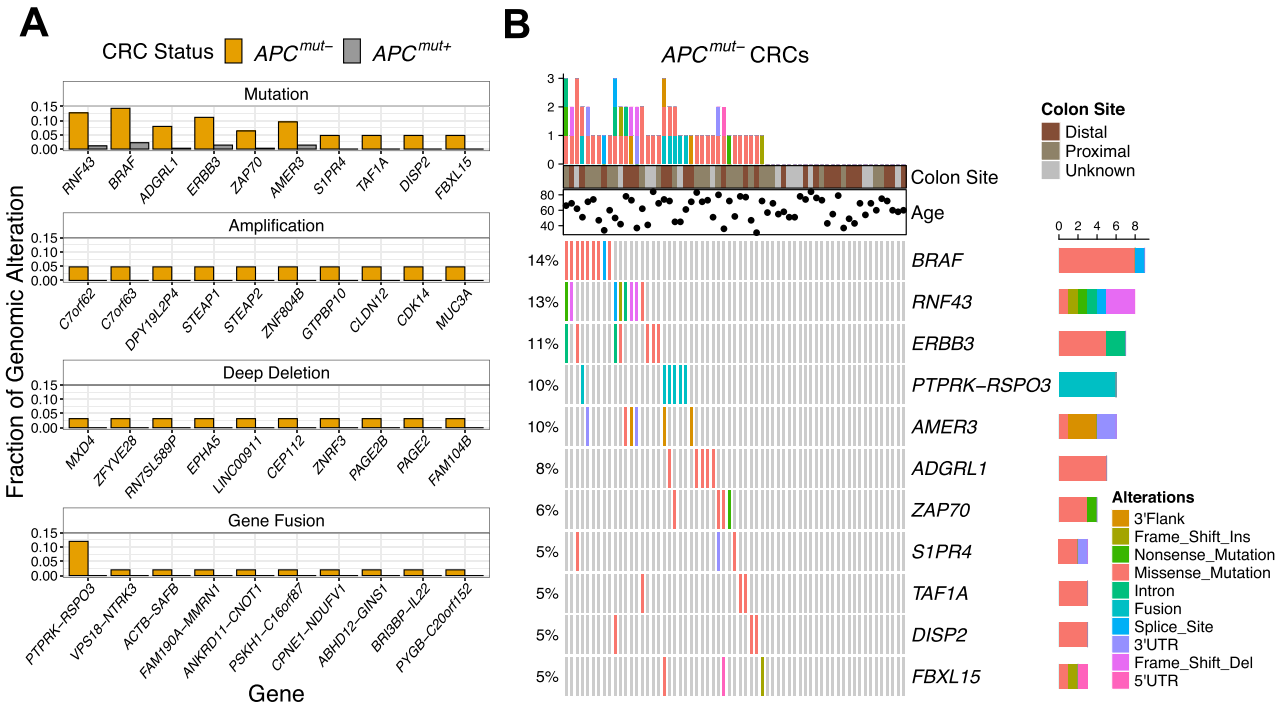
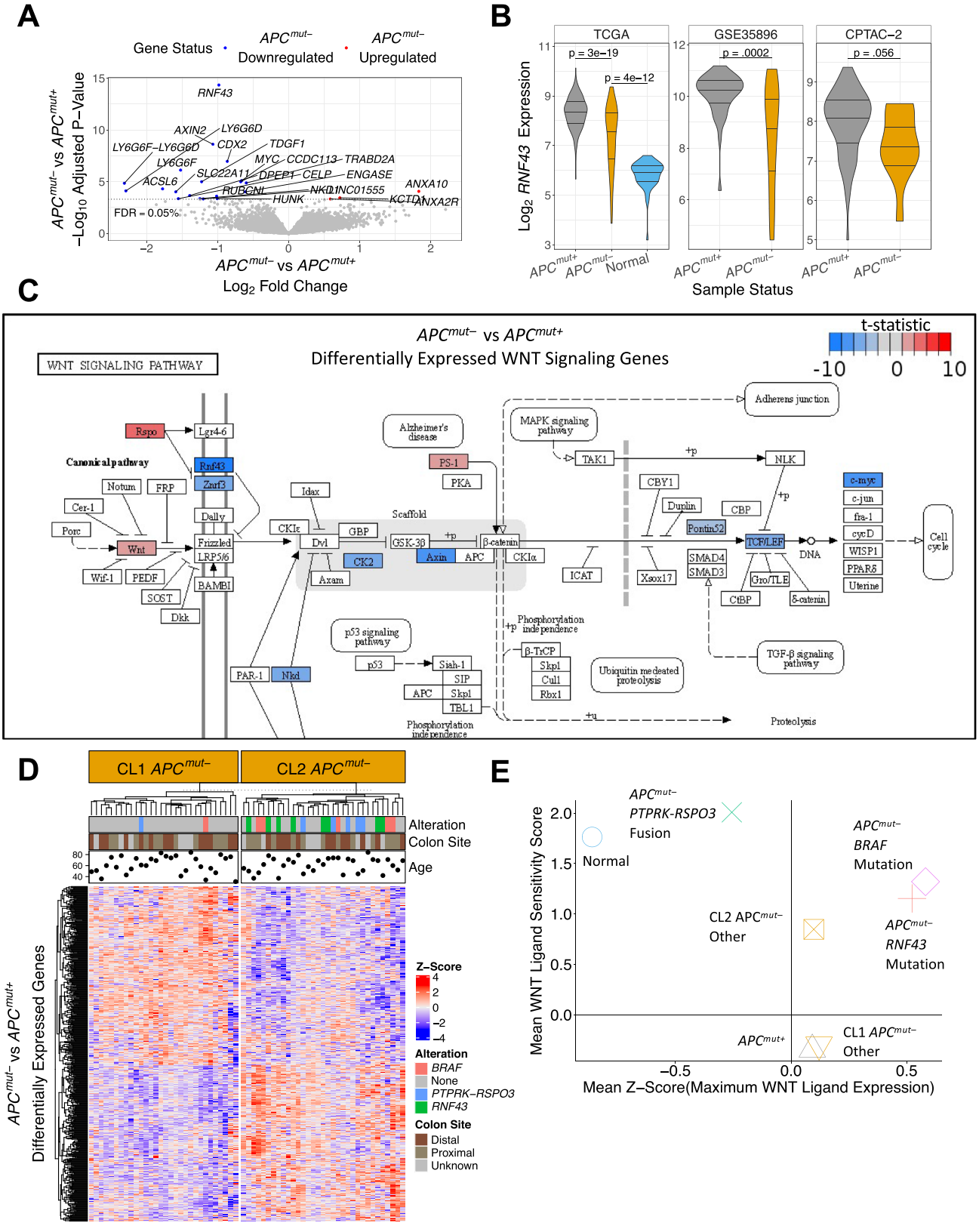


Figure 1. WNT signaling mutations in APC^{mut-} CRCs. (**A**) Fraction of APC^{mut-} CRCs from the TCGA dataset with gene mutations, amplifications, deep deletions, and fusions that were significantly more common in APC^{mut-} in comparison to APC^{mut+} CRCs. The top 10 are shown by p-value ranking, most significant (left) to least (right). (**B**) OncoPrint diagram showing the top 10 statistically significant mutations associated with APC^{mut-} CRCs and the gene fusion *PTPRK-RSPO3* for the 63 APC^{mut-} CRCs.


A
Gene Status
APC^mut– Downregulated
APC^mut– Upregulated
−Log10 Adjusted P−Value
APC^mut− vs APC^mut+
FDR = 0.05%
APC^mut− vs APC^mut+
Log2 Fold Change
RNF43
AXIN2
LY6G6D
CDX2
TDGF1
LY6G6F−LY6G6D
LY6G6F
MYC
CCDC113
TRABD2A
SLC22A11
DPEP1
CELP
ENGASE
ACSL6
RUBCNL
HUNK
NKD1
INC01555
KCTD1
ANXA10
ANXA2R
B
TCGA
GSE35896
CPTAC−2
p = 3e−19
p = 4e−12
p = .0002
p = .056
Log2 RNF43 Expression
APC^mut+
APC^mut−
Normal
Sample Status
C
APC^mut− vs APC^mut+
Differentially Expressed WNT Signaling Genes
t-statistic
-10
0
10
WNT SIGNALING PATHWAY
Canonical pathway
Rspo
Lgr4-6
Rnf43
Znrf3
Notum
FRP
Cer-1
Dally
Porc
Wnt
Frizzled
LRP5/6
Wif-1
PEDF
SOST
Dkk
BAMBI
Idax
CKIε
GBP
Dvl
CK2
Axam
Scaffold
GSK-3β
Axin
APC
β-catenin
CKIα
Alzheimer's disease
PS-1
PKA
MAPK signaling pathway
TAK1
Adherens junction
NLK
CBY1
Duplin
CBP
Pontin52
TCF/LEF
DNA
ICAT
Xsox17
SMAD4
SMAD3
Gro/TLE
CtBP
δ-catenin
c-myc
c-jun
fra-1
cyc D
WISP1
PPARδ
Uterine
Cell cycle
Phosphorylation independence
β-TrCP
Skp1
Cul1
Rbx1
Ubiquitin mediated proteolysis
TGF-β signaling pathway
p53 signaling pathway
p53
Siah-1
SIP
Skp1
TBL1
PAR-1
Nkd
Proteolysis
D
CL1 APC^mut–
CL2 APC^mut–
Alteration
Colon Site
Age
80
60
40
APC^mut− vs APC^mut+
Differentially Expressed Genes
Z–Score
4
2
0
−2
−4
Alteration
BRAF
None
PTPRK–RSPO3
RNF43
Colon Site
Distal
Proximal
Unknown
E
Mean WNT Ligand Sensitivity Score
2.0
1.5
1.0
0.5
0.0
APC^mut– PTPRK-RSPO3 Fusion
Normal
APC^mut– BRAF Mutation
CL2 APC^mut– Other
APC^mut– RNF43 Mutation
APC^mut+
CL1 APC^mut– Other
−0.5
0.0
0.5
Mean Z–Score(Maximum WNT Ligand Expression)

◂**Figure 2.** Enhanced sensitivity to extracellular WNT in APC^{mut-} CRCs. (**A**) Volcano plot representing the results from differential expression analysis between APC^{mut-} and APC^{mut+} CRCs. Labeled points are the genes with an $P_{adj} < 0.0005$. Blue points were downregulated in APC^{mut-} CRCs and red points upregulated. (**B**) Comparison of *RNF43* gene expression in APC^{mut-} CRCs, APC^{mut+} CRCs, and normal colon samples in the TCGA, GSE35896 and CPTAC-2 datasets. Two-sample t-tests with a two tailed p-value were used to test statistical significance. (**C**) Differentially expressed genes ($P_{adj} < 0.05$) between APC^{mut-} and APC^{mut+} CRCs from TCGA were mapped onto the KEGG canonical WNT signaling pathway. Blue labeling represents genes downregulated in APC^{mut-}; red labeling represents upregulated genes. (**D**) Unsupervised clustering analysis of APC^{mut-} CRCs from the TCGA dataset using differentially expressed genes ($P_{adj} < 0.05$). (**E**) Scatter plot showing estimation of activation potential of extracellular WNT signaling. Each point is the mean for individual groups. The y-axis represents a group's apparent sensitivity to extracellular WNT signaling using the WNT ligand sensitivity score. The x-axis represents a group's WNT stimulation potential by quantifying each sample's maximum WNT ligand expression.

tion dataset, 16 out of 56 (29%) CRCs were classified as APC^{mut-} and in the CPTAC-2 dataset, 11 out of 81 (14%) CRCs were classified as APC^{mut-}.

We tested clinical features that might be statistically associated with TCGA APC^{mut-} CRCs (Table 1). In agreement with previous studies[6–8], APC^{mut-} CRCs were diagnosed at a younger age (61.4 in APC^{mut-} vs. 66.4 in APC^{mut+}), and 63% of tumors diagnosed < 50 were APC^{mut-}. APC^{mut-} CRCs were also younger in the CPTAC-2 dataset (61.5 in APC^{mut-} vs. 65.5 in APC^{mut+}), but this observation did not reach statistical significance ($p = 0.24$). (Age of diagnosis was not available for the GSE35896 dataset.) In addition to age, TCGA APC^{mut-} CRCs were more prevalent in Asians ($p = 0.005$), were more likely to be classified as CpG island methylator phenotype (CIMP) high ($p = 0.02$), and were more likely to be diagnosed later than stage one ($p = 0.035$).

WNT signaling mutations in APC^{mut-} CRCs. To identify distinctive somatic mutations, we compared non-silent nucleotide variants, gene amplifications, deep gene deletions, and gene fusions in APC^{mut-} and APC^{mut+} CRCs (Fig. 1A). The top three most statistically different genomic alterations specific to APC^{mut-} CRCs were *PTPRK-RSPO3* gene fusions ($p = 1.3 \times 10^{-5}$), *RNF43* mutations ($p = 4.7 \times 10^{-5}$) and *BRAF* mutations ($p = 1.9 \times 10^{-4}$). These genetic alterations have been identified in CRC previously with evidence for mutual exclusivity with *APC* mutations[27–30]. (The *RNF43* mutation G659Vfs*41, which is associated with MSI CRCs, was not present in the tumors analyzed here as MSI tumors were excluded from this analysis[31]). Eight of nine *BRAF* mutated APC^{mut-} CRCs had the oncogenic V600E *BRAF* mutation. Six of eight *RNF43* mutated APC^{mut-} CRCs had mutations that caused premature protein truncations, whilst one sample had a previously identified missense mutation, R554G. These findings suggested that *BRAF* and *RNF43* mutations are associated with tumor progression in MSS APC^{mut-} CRCs.

Based on the mutated genes in Fig. 1A and the *PTPRK-RSPO3* gene fusion, we found that only 37 out of 63 samples (59%) contained a genomic alteration that distinguished APC^{mut-} from APC^{mut+} CRCs (Fig. 1B). No pairwise combination of genes were statistically mutually exclusive. However, *PTPRK-RSPO3* gene fusions and *RNF43* mutations never co-occurred and were found in 23% of the APC^{mut-} CRCs. After disregarding overlapping genomic alterations, *BRAF* mutations were the next most abundant (10%), followed by mutations in *ADGRL1* (6%), *ERBB3* (5%), and *ZAP70* (5%). Supporting these findings, we found that *BRAF* mutations in the GSE35896 dataset and mutations in *RNF43*, *ERBB3*, and *ZAP70* in the CPTAC-2 dataset were more frequent in APC^{mut-} CRCs than in APC^{mut+} CRCs (Supplementary Fig. 1). (No additional mutation information was provided with the GSE35896 dataset.)

Enhanced sensitivity to extracellular WNT in APC^{mut-} CRCs. Because a distinctive somatic mutational mechanism was not evident in over 40% of APC^{mut-} CRCs, we examined transcriptomics data for further distinguishing molecular characteristics. Strikingly, in differential gene expression analysis of the TCGA dataset, *RNF43* was the most differentially expressed gene between the two tumor groups ($P_{adj} = 4.6 \times 10^{-15}$; Fig. 2A), with a -0.98 $\log_2$ fold decrease in mean expression level in APC^{mut-} CRCs. Consistent with these results, *RNF43* was also down-regulated in APC^{mut-} CRCs in the GSE35896 and CPTAC-2 validation datasets (Fig. 2B). *RNF43* and its family member *ZNRF3* are membrane-bound E3 ubiquitin ligases that actuate the degradation of low-density-lipoprotein-related protein (LRP)-FZD WNT receptors. Binding of R-spondins to leucine-rich repeat-containing G-protein coupled receptors (LGR) leads to sequestration and membrane clearance of RNF43 and ZNRF3 from the cell surface[32–34]. The transcriptional down-regulation of *RNF43* we found in APC^{mut-} CRCs suggested that these tumors may express higher levels of LRP-FZD receptors at the cell surface, and consequently be more responsive to extracellular WNTs.

Because WNT signaling was implicated by these results, we sought to determine the extent to which other factors in the canonical WNT signaling pathway were differentially expressed between APC^{mut-} and APC^{mut+} CRCs (Fig. 2C). Consistent with the results above, we observed that other genes involved in extracellular WNT signaling were dysregulated, namely *RSPO3* and *ZNRF3*. Differences in *RSPO3* and *ZNRF3* mRNA expression showed a similar trend in the validation datasets and were statistically significant in select cases (Supplementary Fig. 2). We did not observe any differential expression of the extracellular WNT regulator genes *LGR4*, *LGR5*, *LGR6,* or LRP-FZD receptors. The fact that *LGR4-6* were not differentially expressed between APC^{mut-} and APC^{mut+} CRCs was consistent with the finding that *RSPO3* does not

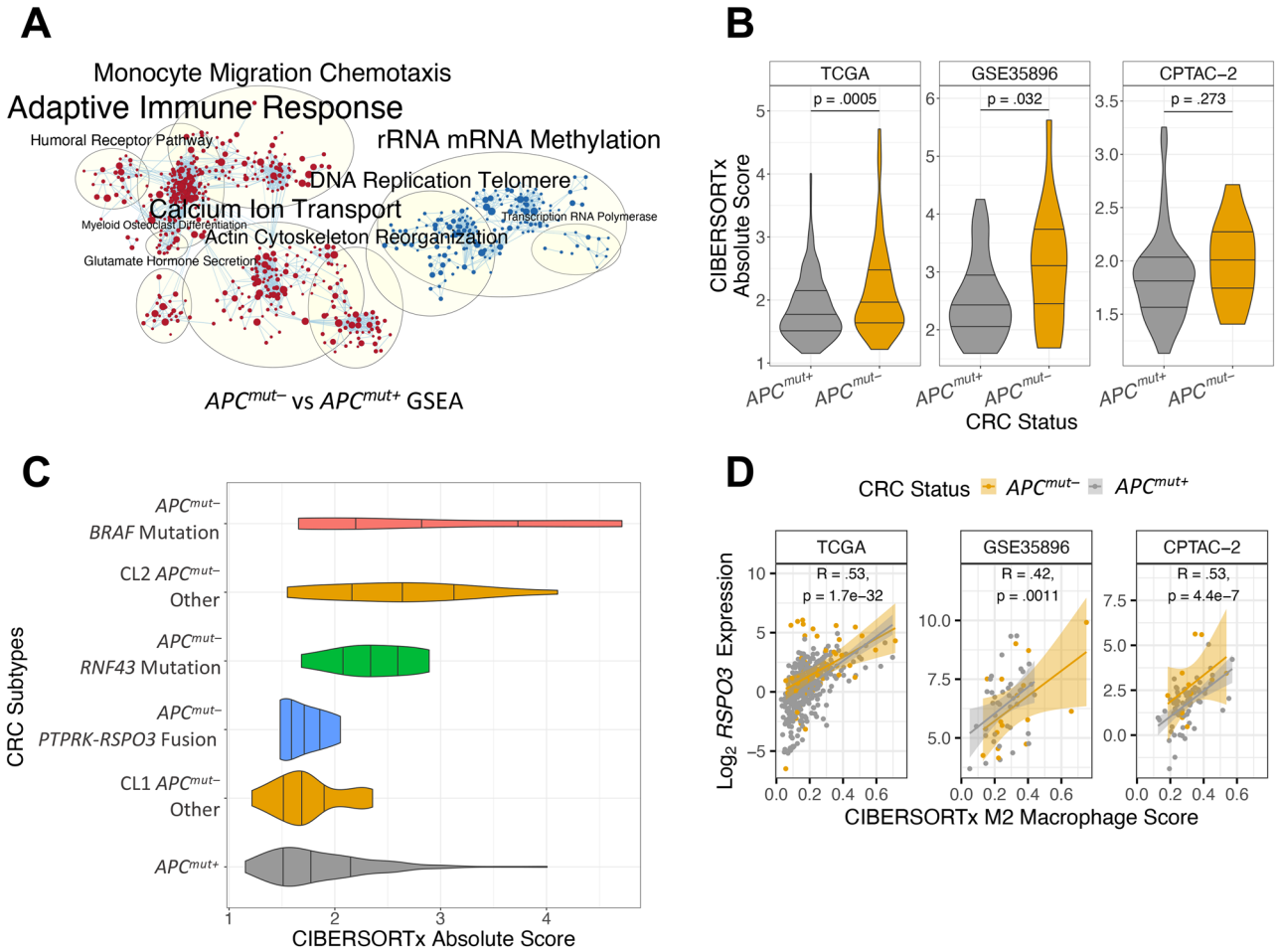


Figure 3. *APC*$^{mut-}$ CRCs associated with immune infiltration. (**A**) GSEA results of differential gene expression analysis of *APC*$^{mut-}$ versus *APC*$^{mut+}$ CRCs from the TCGA dataset. Red clusters represent GO terms enriched among upregulated genes in *APC*$^{mut-}$ CRCs and blue clusters correspond to down-regulated processes. (**B**) CIBERSORTx absolute score in CRCs from the TCGA, GSE35896 and CPTAC-2 datasets. Two-sample t-tests with a two-tailed p-value were used to test statistical significance. (**C**) Violin plot of CIBERSORTx absolute score across subtypes of *APC*$^{mut-}$ CRCs. (**D**) Expression of *RSPO3* in *APC*$^{mut-}$ and *APC*$^{mut+}$ CRCs plotted against their individual M2 macrophage scores identified from the CIBERSORTx algorithm. Pearson correlation was performed to determine statistical significance.

require interaction with LGRs to potentiate WNT signaling[35] and LRP-FZD receptor levels are regulated post-transcriptionally[36]. When we compared gene expression of *APC*$^{mut-}$ and *APC*$^{mut+}$ CRCs to normal samples and mapped genes onto the canonical WNT signaling pathway, changes in gene expression in WNT signaling were similar between these two tumor types (Supplementary Fig. 3). These results suggested that both types of CRCs exploit changes in WNT signaling. However, based on the mutation and expression data, *APC*$^{mut-}$ CRCs appear to favor dysregulation of genes involved in response to extracellular WNT signaling, whereas *APC*$^{mut+}$ CRCs are stuck in the "on" state and are WNT signal-transduction incompetent.

To determine the fraction of *APC*$^{mut-}$ CRCs that operate via enhanced sensitivity of extra-cellular WNT, we performed unsupervised hierarchical clustering using all differentially expressed genes ($P_{adj} < 0.05$) between *APC*$^{mut-}$ and *APC*$^{mut+}$ CRCs in the TCGA dataset (Fig. 2D). *APC*$^{mut-}$ CRCs clustered into two prominent groups, referred to here as Cluster 1 (CL1) and Cluster 2 (CL2). Most *APC*$^{mut-}$ CRCs with a *PTPRK-RSPO3* fusion, *BRAF* mutation, or *RNF43* mutation were in CL2. To characterize the expression profiles of *APC*$^{mut-}$ CRCs in the context of extracellular WNT signaling, we computed a summarized score defined as *RSPO3* expression minus the sum of *RNF43* and *ZNRF3* expression. This score represents a theoretical WNT ligand sensitivity (WNT_{LS}) based on the known function of *RSPO3* in increasing ligand sensitivity, and *RNF43* and *ZNRF3* in decreasing ligand sensitivity[33,34]. We examined how the WNT_{LS} score tracked with maximum WNT ligand expression (Fig. 2E; see Methods for more details). Consistent with our expectation, *APC*$^{mut-}$ CRCs with *RNF43* mutations had higher WNT_{LS} scores than *APC*$^{mut+}$ CRCs and higher maximum WNT expression than normals. Interestingly, *APC*$^{mut-}$ CRCs with *PTPRK-RSPO3* fusions had the highest WNT_{LS} score, but had the lowest expression of WNT ligands compared to other CRCs. Inconsistencies in how CRCs with *PTPRK-RSPO3* fusions and CRCs

with *RNF43* mutations enhance their sensitivity to extracellular WNT signaling may be due to different selective pressures during cancer evolution.

APC^{mut-} CRCs with *BRAF* mutations also exhibited higher WNT_{LS} and higher WNT ligand expression, similar to *APC*$^{mut-}$ CRCs with *RNF43* mutations. Importantly, *APC*$^{mut-}$ CRCs from CL2 that did not have *BRAF* mutations, *RNF43* mutations, or *PTPRK-RSPO3* fusions exhibited higher WNT_{LS} scores compared to *APC*$^{mut+}$ CRCs. In contrast, *APC*$^{mut-}$ CRCs from CL1 exhibited WNT_{LS} scores similar to *APC*$^{mut+}$ CRCs. *APC*$^{mut-}$ CRCs from the GSE35896 and CPTAC-2 datasets also clustered into two groups with high and low WNT_{LS} scores (Supplementary Fig. 4). Given the importance of WNT signaling in CRC, these results suggest that other WNT-related mechanisms drive CL1 *APC*$^{mut-}$ CRCs. By transcriptomic analysis, CL1 *APC*$^{mut-}$ CRCs were practically indistinguishable from *APC*$^{mut+}$ CRCs; however, GSEA showed enrichment of oxidative phosphorylation genes (Supplementary Figs. 5 and 6), implicating mitochondrial activation in CL1 *APC*$^{mut-}$ tumorigenesis. These results were supported by data from the DepMap CRISPR screen that demonstrated dependence of *APC*$^{mut-}$ CRC cell lines on oxidative phosphorylation complexes in the mitochondria (Supplementary Fig. 5E).

APC^{mut-} CRCs associated with immune infiltration. GSEA analysis showed that GO terms related to the adaptive immune response were upregulated in *APC*$^{mut-}$ compared to *APC*$^{mut+}$ CRCs (Fig. 3A). To further investigate immune system involvement in *APC*$^{mut-}$ CRCs, we employed the bulk tissue deconvolution method CIBERSORTx[22]. In agreement with the GSEA results, the CIBERSORTx absolute score was increased in *APC*$^{mut-}$ compared to *APC*$^{mut+}$ CRCs in all three CRC datasets (Fig. 3B). The CIBERSORTx absolute score was highest in *APC*$^{mut-}$ CRCs with *BRAF* or *RNF43* mutations and CL2 *APC*$^{mut-}$ CRCs without mutations (Fig. 3C). Because these *APC*$^{mut-}$ CRCs had more infiltrating immune cells than those with *PTPRK-RSPO3* fusions, we tested whether any of the 22 immune cell types were associated with expression of WNT agonist ligand *RSPO3* (Fig. 3D). We found that M2 macrophages had the strongest positive Pearson correlation with *RSPO3* expression. M2 macrophages and *RSPO3* expression were also significantly correlated in the GSE35896 and the CPTAC-2 datasets. Macrophage expression of *RSPO3* was shown in a study of patients with pulmonary fibrosis[37].

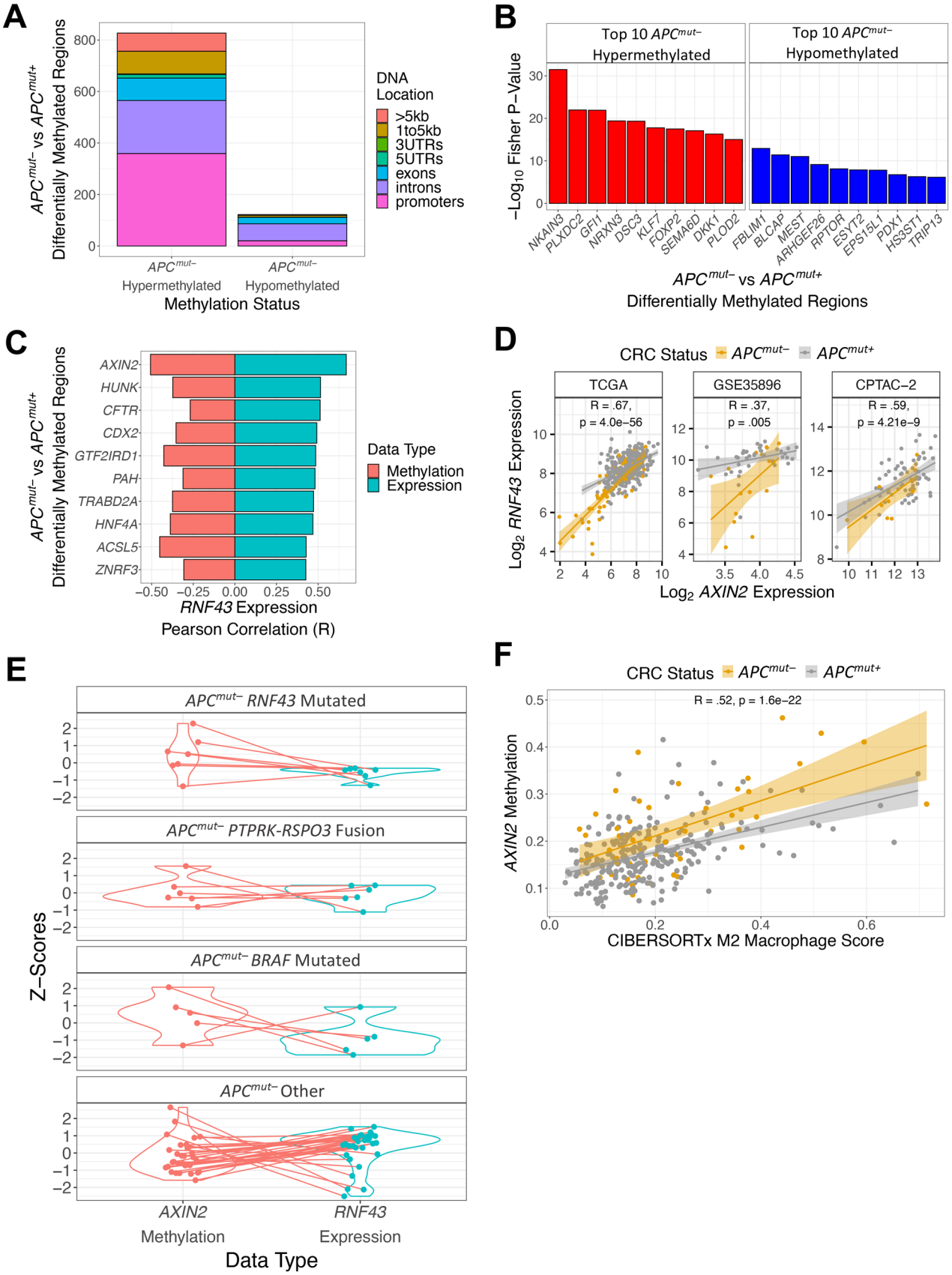
APC^{mut-} CRCs have higher _AXIN2_ methylation. Because we found an association between *APC*$^{mut-}$ CRCs and CIMP-high previously[7], we identified differentially methylated regions (DMRs) between *APC*$^{mut-}$ and *APC*$^{mut+}$ CRCs. *APC*$^{mut-}$ CRCs were globally more hypermethylated than *APC*$^{mut+}$ CRCs, with a particular excess in promoter regions (Fig. 4A). Comparing the top ten hypermethylated and hypomethylated DMRs, we did not observe the same statistically significant genes as we did in the mutation and expression analyses (Fig. 4B). However, when we tested correlation of *RNF43* expression with DNA methylation levels of DMRs and with RNA expression, we found that methylation and gene expression of *AXIN2* had the highest correlations (Fig. 4C). *RNF43* gene expression was also significantly correlated with *AXIN2* expression in the GSE35896 and CPTAC-2 datasets (Fig. 4D; methylation data was not available in these datasets). Increased *AXIN2* DNA methylation was associated with decreased *RNF43* expression in a subset of *APC*$^{mut-}$ CRCs that did not have one of the common somatic mutations (Fig. 4E). Similar to our findings with *RSPO3* expression, we found that M2 macrophages correlated most with *AXIN2* DNA methylation (Fig. 4F).

_AP2M1_ gene expression associated with earlier onset in _APC^{mut-}_ CRCs. Age of onset was not different in CL1 and CL2 *APC*$^{mut-}$ CRCs (Fig. 5A). To identify gene expression changes linked to earlier age of onset in *APC*$^{mut-}$ CRCs, we separated *APC*$^{mut-}$ CRCs into two groups based on the median expression of each gene and performed a logrank test between these two groups, using the age at diagnosis as the event variable. Expression of *AP2M1* best distinguished earlier onset *APC*$^{mut-}$ CRCs from later onset *APC*$^{mut-}$ CRCs (Fig. 5B), and higher *AP2M1* expression was associated with earlier onset in *APC*$^{mut-}$ relative to *APC*$^{mut+}$ CRCs (Fig. 5C).

Discussion

Most CRCs are initiated by somatic mutation of the gene *APC,* leading to ligand-independent, constitutive activity of the WNT pathway. In the present study, we found two alternate ways in which *APC*$^{mut-}$ CRCs may activate the WNT pathway. *APC*$^{mut-}$ tumors clustered into two groups according to their transcriptomic profiles (Fig. 5D). One cluster (CL2) exhibited a variety of molecular alterations that were consistent with the hypothesis that these tumors have enhanced sensitivity to extracellular WNT ligands. In particular, the most significant change was downregulation of *RNF43*, which is expected to result in increased levels of WNT receptors and greater sensitivity to extracellular WNTs. *AXIN2* methylation was highly correlated with *RNF43* downregulation. AXIN2 and RNF43 are negative regulators of WNT signaling that are transcriptionally activated by nuclear β-catenin, consistent with the notion that epigenetic silencing of negative regulators plays a critical role in tumor formation in ligand-dependent, *APC*$^{mut-}$ CRCs. Similarly, *PTPRK-RSPO3* gene fusions drive R-spondin signaling, which is also expected to reduce RNF43 levels at the cell surface, upregulate WNT receptors, and enhance sensitivity to extracellular WNTs. We defined a WNT ligand sensitivity score to quantify this signature of extracellular WNT signaling in a sample-specific fashion and found a high WNT_{LS} score was associated with CL2 *APC*$^{mut-}$ CRCs in multiple independent datasets.

Germline mutations in *RNF43* have been previously associated with serrated polyposis families, and somatic mutations in *RNF43* and *BRAF* have been associated with sporadic serrated adenomas[38]. In a preliminary analysis, we found that CL2 *APC*$^{mut-}$ CRCs expression profiles appear to be more similar


A
APC^mut– vs APC^mut+ Differentially Methylated Regions
800
600
400
200
0
DNA Location
>5kb
1to5kb
3UTRs
5UTRs
exons
introns
promoters
APC^mut– Hypermethylated
APC^mut– Hypomethylated
Methylation Status
B
Top 10 APC^mut– Hypermethylated
Top 10 APC^mut– Hypomethylated
−Log10 Fisher P−Value
30
20
10
0
NKAIN3
PLXDC2
GFI1
NRXN3
DSC3
KLF7
FOXP2
SEMA6D
DKK1
PLOD2
FBLIM1
BLCAP
MEST
ARHGEF26
RPTOR
ESYT2
EPS15L1
PDX1
HS3ST1
TRIP13
APC^mut– vs APC^mut+ Differentially Methylated Regions
C
APC^mut– vs APC^mut+ Differentially Methylated Regions
AXIN2
HUNK
CFTR
CDX2
GTF2IRD1
PAH
TRABD2A
HNF4A
ACSL5
ZNRF3
−0.50 −0.25 0.00 0.25 0.50
RNF43 Expression Pearson Correlation (R)
Data Type
Methylation
Expression
D
CRC Status APC^mut– APC^mut+
TCGA
R = .67, p = 4.0e−56
GSE35896
R = .37, p = .005
CPTAC−2
R = .59, p = 4.21e−9
Log2 RNF43 Expression
Log2 AXIN2 Expression
E
APC^mut– RNF43 Mutated
APC^mut– PTPRK-RSPO3 Fusion
APC^mut– BRAF Mutated
APC^mut– Other
Z−Scores
AXIN2 Methylation
RNF43 Expression
Data Type
F
CRC Status APC^mut– APC^mut+
R = .52, p = 1.6e−22
AXIN2 Methylation
CIBERSORTx M2 Macrophage Score

◂**Figure 4.** *APC*$^{mut-}$ CRCs have higher *AXIN2* methylation. (**A**) Bar plot comparing total number of hyper-methylated and hypo-methylated differentially methylated regions (DMRs) between *APC*$^{mut-}$ and *APC*$^{mut+}$ CRCs from the TCGA dataset. (**B**) Top 10 *APC*$^{mut-}$ hypermethylated and hypomethylated DMRs between *APC*$^{mut-}$ and *APC*$^{mut+}$ CRCs from TCGA. Red bars represent *APC*$^{mut-}$ CRC differentially hypermethylated genes and blue bars represent *APC*$^{mut-}$ CRC differentially hypomethylated genes. (**C**) Bar plot representing DMRs with strongest correlations with *RNF43*. Blue bars represent the top 10 DMRs with the highest Pearson gene expression correlation with *RNF43* gene expression. Red bars represent the Pearson correlation between the average differentially methylated beta values and *RNF43* expression for these differentially methylated regions. (**D**) Scatter plots of *RNF43* expression and *AXIN2* expression of both *APC*$^{mut-}$ and *APC*$^{mut+}$ CRCs in the TCGA, GSE35896, and CPTAC-2 datasets. Pearson correlation was performed to determine statistical significance. (**E**) Matched comparison between Z-normalized *AXIN2* average beta values and Z-normalized *RNF43* expression of *APC*$^{mut-}$ CRCs. (**F**) Scatter plot of *AXIN2* average beta values and the CIBERSORTx M2 macrophage score of *APC*$^{mut-}$ and *APC*$^{mut+}$ CRCs from the TCGA dataset. Pearson correlation was performed to measure statistical significance.

to serrated adenomas than CL1 *APC*$^{mut-}$ CRCs and *APC*$^{mut+}$ CRCs, according to two published gene signatures[39,40] (data not shown), but these results need further investigation.

We also found that CL2 *APC*$^{mut-}$ CRCs have a higher level of immune infiltration compared to *APC*$^{mut+}$ and CL1 *APC*$^{mut-}$ CRCs, especially in *APC*$^{mut-}$ CRCs that had *RNF43* or *BRAF* mutations. M2 macrophages had the strongest association with potentiating WNT signaling through its significant correlations with *RSPO3* expression and *AXIN2* DNA methylation. Previous studies have shown that macrophages have the capability to express *RSPO3* and stimulate WNT signaling in response to tissue damage[41,42]. The association of CL2 *APC*$^{mut-}$ CRCs with M2 macrophages suggests the etiology of this cancer subtype is tied to chronic tissue stress and inflammation that eventually favors a clone with hypersensitivity to WNT. We suggest that CL2 *APC*$^{mut-}$ CRCs may be sensitive to porcupine inhibitors or anti-WNT/anti-DKK1 biologics. We note that *AXIN2* methylation has been previously identified in *APC*$^{mut-}$ CRCs as a potential biomarker for ligand-dependent tumors that would respond to anti-WNT-based therapies such as porcupine inhibitors[43–45].

In contrast, the other cluster (CL1) of *APC*$^{mut-}$ CRCs was associated with low WNT_{LS} score and may be dependent on enhanced mitochondrial activation. *APC*$^{mut-}$ CRC cell lines from the DepMap database had a strong dependency on mitochondrial activation relative to *APC*$^{mut+}$ CRC cell lines. We are cautious in interpreting these data, because the observed effectiveness of mitochondrial disruption of the *APC*$^{mut-}$ CRC cell lines may be due to the absence of immune cells in vitro. One potential reason why some *APC*$^{mut-}$ CRCs become dependent on enhanced mitochondrial activation is because mitochondria can stimulate the WNT pathway independently of WNT ligands[46]. Moreover, intestinal epithelial cell-specific knockout of TFAM, a transcription factor required for replication of mitochondria DNA, drastically reduced tumor formation in *APC*$^{min/+}$ mouse models[47]. Therefore, we suggest that enhanced activation of mitochondria is a second, independent mechanism by which *APC*$^{mut-}$ CRCs exploit WNT signaling in tumor progression. These findings also suggest that mitochondria inhibitors may be a promising therapeutic option for CL1 *APC*$^{mut-}$ CRCs.

Although *APC*$^{mut-}$ tumors overall exhibit a lower age of onset than APC^{mut+} tumors, we found no difference in age of onset between CL1 and CL2, suggesting that both extracellular WNT sensitivity and mitochondrial activation contribute to the incidence of early-onset CRC. We performed a *APC*$^{mut-}$-wide analysis to determine what gene expression feature was most associated with age of onset and found that earlier-onset *APC*$^{mut-}$ CRCs had increased expression of *AP2M1*. *AP2M1* plays an important role in clathrin-mediated endocytosis[48]. A recent study showed that when insulin binds to an insulin receptor, *IRS1* and *IRS2* recruit *AP2M1* to initiate insulin receptor endocytosis[49]. Thus, an increase of *AP2M1* may suggest increased insulin signaling. Importantly, insulin can activate both the PI3K pathway and the MAPK pathway, which may in turn play a role in enhancing both mitochondrial activation and immune infiltration, thus contributing to driving both CL1 and CL2 subtypes of *APC*$^{mut-}$ CRC[50–52]. Other studies have found that individuals with type two diabetes are at a greater risk for early-onset CRC[53,54].

Early-onset CRC is a rapidly advancing public health emergency, and it is associated with a lack of mutation in *APC*. Our comprehensive genomic analysis has uncovered two classes of *APC*$^{mut-}$ CRCs, one which potentiates WNT signaling through sensitivity to extracellular signaling, and the other which exhibits mitochondrial activation. Future research should test the effect of anti-WNT biologics and mitochondrial inhibitors in organoid

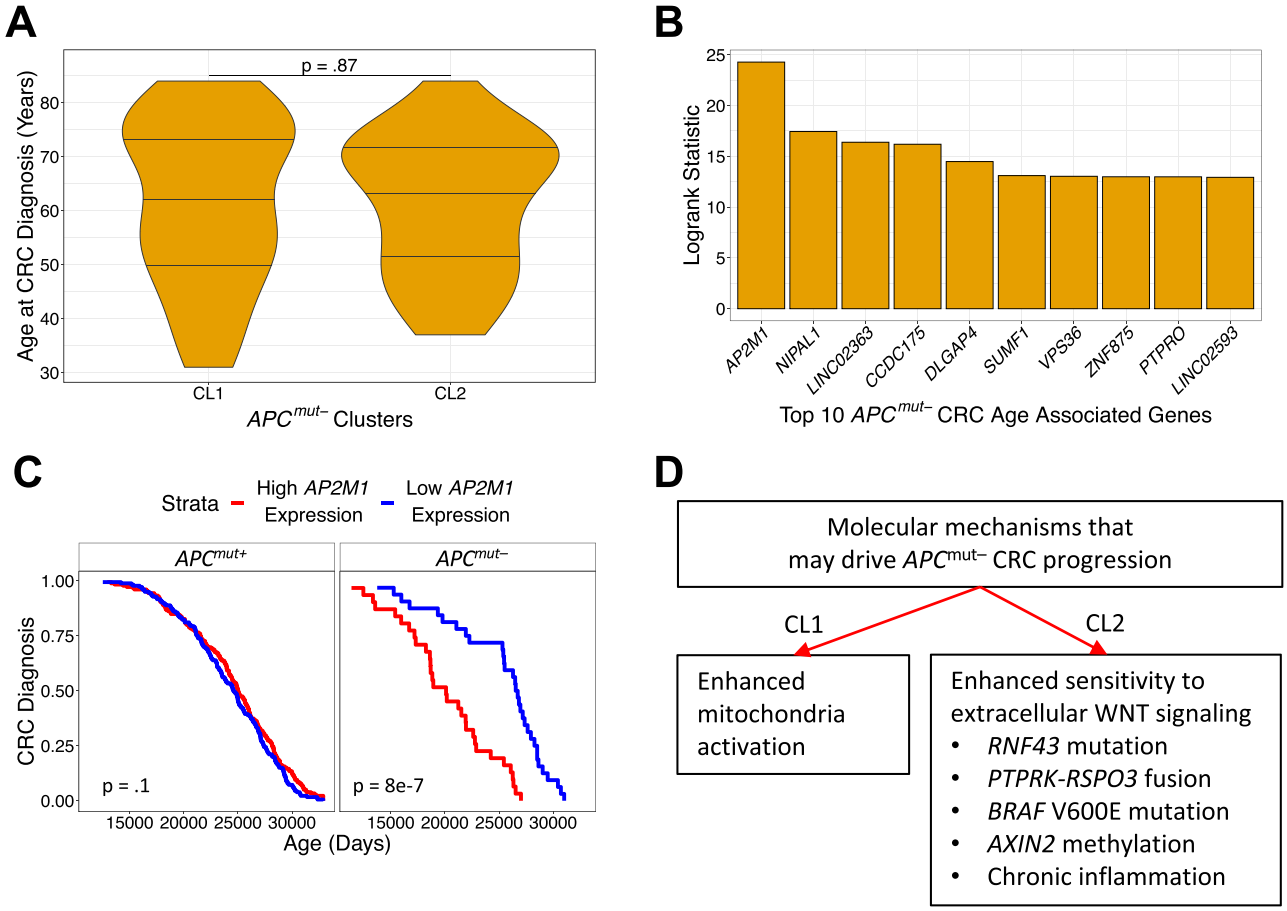


Figure 5. *AP2M1* gene expression associated with earlier-onset in APC^{mut-} CRCs. (**A**) A comparison of age between APC^{mut-} clusters identified from Fig. 2D. A two-sample t-test with a two-tailed p-value was used to determine statistical significance. (**B**) Top 10 statistically significant genes based on a logrank test whose median gene expression best separates age of CRC diagnosis of APC^{mut-} CRCs from the TCGA dataset. (**C**) Kaplan–Meier plot representing association between the age at CRC diagnosis and median separation of *AP2M1* expression in APC^{mut-} and APC^{mut+} CRCs from the TCGA dataset. (**D**) Flowchart of two molecular mechanisms that may be involved in the development of APC^{mut-} CRC.

models and in vivo and compare the efficacy of *AXIN2* methylation and WNT ligand sensitivity score in identifying anti-WNT sensitive tumors.

Data availability

Data analyzed in this study can be found in the Genomic Data Commons (https://gdc.cancer.gov); Gene Expression Omnibus (GSE35896; https://www.ncbi.nlm.nih.gov/geo/); cBioPortal (CPTAC-2; https://www.cbioportal.org/); and DepMap (https://depmap.org/portal/). All analytic methods and study materials are available to other researchers through supplemental materials and in the methods section.

Received: 8 September 2021; Accepted: 18 November 2021
Published online: 06 December 2021

References

1. Siegel, R. L. *et al.* Colorectal cancer statistics, 2020. *CA Cancer J. Clin.* **70**, 145–164 (2020).
2. Li, D. Recent advances in colorectal cancer screening. *Chronic Dis. Transl. Med.* **4**, 139–147 (2018).
3. Mauri, G. *et al.* Early-onset colorectal cancer in young individuals. *Mol. Oncol.* **13**, 109–131 (2019).
4. Pearlman, R. *et al.* Prevalence and spectrum of germline cancer susceptibility gene mutations among patients with early-onset colorectal cancer. *JAMA Oncol.* **3**, 464–471 (2017).
5. Hofseth, L. J. *et al.* Early-onset colorectal cancer: Initial clues and current views. *Nat. Rev. Gastroenterol. Hepatol.* **17**, 352–364 (2020).
6. Lieu, C. H. *et al.* Comprehensive genomic landscapes in early and later onset colorectal cancer. *Clin. Cancer Res.* **25**, 5852–5858 (2019).
7. Xicola, R. M. *et al.* Lack of APC somatic mutation is associated with early-onset colorectal cancer in African Americans. *Carcinogenesis* **39**, 1331–1341 (2018).
8. Willauer, A. N. *et al.* Clinical and molecular characterization of early-onset colorectal cancer. *Cancer* **125**, 2002–2010 (2019).
9. Fodde, R. The APC gene in colorectal cancer. *Eur. J. Cancer* **38**, 867–871 (2002).

10. Schneikert, J. & Behrens, J. The canonical Wnt signalling pathway and its APC partner in colon cancer development. *Gut* **56**, 417–425 (2007).
11. Lochhead, P. *et al.* Microsatellite instability and BRAF mutation testing in colorectal cancer prognostication. *J. Natl. Cancer Inst.* **105**, 1151–1156 (2013).
12. Colaprico, A. *et al.* TCGAbiolinks: An R/bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).
13. Hu, X. *et al.* TumorFusions: An integrative resource for cancer-associated transcript fusions. *Nucleic Acids Res.* **46**, D1144–D1149 (2018).
14. Kim, S. & Jeong, S. Mutation hotspots in the beta-catenin gene: Lessons from the human cancer genome databases. *Mol. Cells* **42**, 8–16 (2019).
15. Leiserson, M. D., Wu, H. T., Vandin, F. & Raphael, B. J. CoMEt: A statistical approach to identify combinations of mutually exclusive alterations in cancer. *Genome Biol.* **16**, 160 (2015).
16. Vasaikar, S. *et al.* Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell* **177**, 1035–1049 (2019).
17. Schlicker, A. *et al.* Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines. *BMC Med. Genom.* **5**, 66 (2012).
18. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
19. Luo, W., Pant, G., Bhavnasi, Y. K., Blanchard, S. G. Jr. & Brouwer, C. Pathview Web: User friendly pathway visualization and data integration. *Nucleic Acids Res.* **45**, W501–W508 (2017).
20. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
21. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* **5**, e13984 (2010).
22. Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).
23. Fortin, J. P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* **15**, 503 (2014).
24. Peters, T. J. *et al.* De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* **8**, 6 (2015).
25. Cavalcante, R. G. & Sartor, M. A. Annotatr: Genomic regions in context. *Bioinformatics* **33**, 2381–2383 (2017).
26. Ghandi, M. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
27. Seshagiri, S. *et al.* Recurrent R-spondin fusions in colon cancer. *Nature* **488**, 660–664 (2012).
28. Giannakis, M. *et al.* RNF43 is frequently mutated in colorectal and endometrial cancers. *Nat. Genet.* **46**, 1264–1266 (2014).
29. Fennell, L. J. *et al.* APC mutation marks an aggressive subtype of BRAF mutant colorectal cancers. *Cancers (Basel)* **12**, 1171 (2020).
30. Li, S. *et al.* Commonly observed RNF43 mutations retain functionality in attenuating Wnt/beta-catenin signaling and unlikely confer Wnt-dependency onto colorectal cancers. *Oncogene* **39**, 3458–3472 (2020).
31. Tu, J. *et al.* The most common RNF43 mutant G659Vfs*41 is fully functional in inhibiting Wnt signaling and unlikely to play a role in tumorigenesis. *Sci. Rep.* **9**, 18557 (2019).
32. de Lau, W., Peng, W. C., Gros, P. & Clevers, H. The R-spondin/Lgr5/Rnf43 module: Regulator of Wnt signal strength. *Genes Dev.* **28**, 305–316 (2014).
33. Hao, H. X. *et al.* ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. *Nature* **485**, 195–200 (2012).
34. Koo, B. K. *et al.* Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. *Nature* **488**, 665–669 (2012).
35. Lebensohn, A. M. & Rohatgi, R. R-spondins can potentiate WNT signaling without LGRs. *Elife* **7**, e33126 (2018).
36. Hao, H. X., Jiang, X. & Cong, F. Control of Wnt receptor turnover by R-spondin-ZNRF3/RNF43 signaling module and its dysregulation in cancer. *Cancers (Basel)* **8**, 54 (2016).
37. Reyfman, P. A. *et al.* Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 1517–1536 (2019).
38. Yan, H. H. N. *et al.* RNF43 germline and somatic mutation in serrated neoplasia pathway and its association with BRAF mutation. *Gut* **66**, 1645–1656 (2017).
39. Kanth, P. *et al.* Gene signature in sessile serrated polyps identifies colon cancer subtype. *Cancer Prev. Res. (Phila Pa)* **9**, 456–465 (2016).
40. Rahmatallah, Y. *et al.* Platform-independent gene expression signature differentiates sessile serrated adenomas/polyps and hyperplastic polyps of the colon. *BMC Med. Genom.* **10**, 81 (2017).
41. Vannella, K. M. & Wynn, T. A. Mechanisms of organ injury and repair by macrophages. *Annu. Rev. Physiol.* **79**, 593–617 (2017).
42. Cosin-Roger, J., Ortiz-Masia, M. D. & Barrachina, M. D. Macrophages as an emerging source of Wnt ligands: Relevance in mucosal integrity. *Front. Immunol.* **10**, 2297 (2019).
43. Kleeman, S. O. *et al.* Exploiting differential Wnt target gene expression to generate a molecular biomarker for colorectal cancer stratification. *Gut* **69**, 1092–1103 (2020).
44. Madan, B. *et al.* Wnt addiction of genetically defined cancers reversed by PORCN inhibition. *Oncogene* **35**, 2197–2207 (2016).
45. Storm, E. E. *et al.* Targeting PTPRK-RSPO3 colon tumours promotes differentiation and loss of stem-cell function. *Nature* **529**, 97–100 (2016).
46. Delgado-Deida, Y., Alula, K. M. & Theiss, A. L. The influence of mitochondrial-directed regulation of Wnt signaling on tumorigenesis. *Gastroenterol. Rep. (Oxf.)* **8**, 215–223 (2020).
47. Wen, Y. A. *et al.* The mitochondrial retrograde signaling regulates Wnt signaling to promote tumorigenesis in colon cancer. *Cell Death Differ.* **26**, 1955–1969 (2019).
48. Jackson, L. P. *et al.* A large-scale conformational change couples membrane recruitment to cargo binding in the AP2 clathrin adaptor complex. *Cell* **141**, 1220–1229 (2010).
49. Choi, E. *et al.* Mitotic regulators and the SHP2-MAPK pathway promote IR endocytosis and feedback regulation of insulin signaling. *Nat. Commun.* **10**, 1473 (2019).
50. Ren, Z. *et al.* Insulin promotes mitochondrial respiration and survival through PI3K/AKT/GSK3 pathway in human embryonic stem cells. *Stem Cell Rep.* **15**, 1362–1376 (2020).
51. Naderi, N., Zamanian Azodi, M., Daskar Abkenar, E., Shahidi Dadras, M. & Talaei, R. Insulin dysregulation plays a critical role in colon inflammation: A bioinformatics approach. *Gastroenterol. Hepatol. Bed Bench* **11**, 85–91 (2018).
52. Huang, P., Han, J. & Hui, L. MAPK signaling in inflammation-associated cancer development. *Protein Cell* **1**, 218–226 (2010).
53. Ali Khan, U. *et al.* Personal history of diabetes as important as family history of colorectal cancer for risk of colorectal cancer: A nationwide cohort study. *Am. J. Gastroenterol.* **115**, 1103–1109 (2020).
54. Mikaeel, R. R. *et al.* Young-onset colorectal cancer is associated with a personal history of type 2 diabetes. *Asia Pac. J. Clin. Oncol.* **17**, 131–138 (2021).

Acknowledgements
Special thanks to Ellen Kaye and the Maltz Foundation for their generous support of University of Arizona Cancer Biology GIDP student AG.

Author contributions
Study concept and design: A.G., R.M.X., X.L., N.E., M.P. Acquisition of data: A.G. Analysis and interpretation of data: A.G., R.M.X., V.N., J.L., C.T., B.S., X.L., N.E., M.P. Drafting of manuscript: A.G., M.P. Critical revision of the manuscript for intellectual content: A.G., R.M.X., C.T., B.S., X.L., N.E., M.P. Statistical analysis: A.G., M.P. Obtained funding: X.L., N.E.

Funding
This work was supported by Colorectal Cancer Alliance's Chris4Life Grant; NIH/NCI CA242914; NIH/NCI P30 CA023074.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-02806-x.

Correspondence and requests for materials should be addressed to M.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.