HHS Public Access

Author manuscript

Cancer. Author manuscript; available in PMC 2019 March 01.

Published in final edited form as:

Cancer. 2018 March 01; 124(5): 1070–1082. doi:10.1002/cncr.31136.

A Mutational Comparison of Adult and Adolescent and Young Adult (AYA) Colon Cancer

James V. Tricoli, Ph.D.[1], **Lisa A. Boardman, M.D.**[2], **Rajesh Patidar, Ph.D.**[3], **Sivasish Sindiri, Ph.D.**[3], **Jin S. Jang, Ph.D.**[2], **William D. Walsh, B.A.**[4], **Paul M. McGregor III, B.A.**[5], **Corinne E. Camalier, M.A.**[4], **Michele G. Mehaffey, M.A.**[4], **Wayne L. Furman, M.D.**[6], **Armita Bahrami, M.D.**[6], **P. Mickey Williams, Ph.D.**[4], **Chih-Jian Lih, Ph.D.**[4], **Barbara A. Conley, M.D.**[1], and **Javed Khan, M.D.**[3]

[1]Division of Cancer Treatment and Diagnosis, National Cancer Institute, Rockville, MD

[2]Mayo Clinic, Rochester, MN

[3]Center for Cancer Research, National Cancer Institute, Bethesda, MD

[4]Molecular Characterization and Assay Development Laboratory, Leidos, Frederick, MD

[5]Personal Genome, Towson, MD

[6]St. Jude Children's Research Hospital, Memphis, TN

Abstract

Background—It is possible that the relative lack of progress in treatment outcome among the adolescent and young adult (AYA) group of cancer patients is due to a difference in disease biology compared to the corresponding diseases in younger and older individuals. There is evidence that colon cancer is more aggressive, and has a poorer prognosis in AYA patients than that observed in older adult patients.

Methods—In order to further understand the molecular basis for this difference we conducted whole exome sequencing (WES) on a cohort of 30 adult, 30 AYA, and 2 pediatric colon cancers.

Results—We detected a statistically significant difference in mutational frequency between AYA and adult samples in 43 genes, including *ROBO1*, *MYCBP2*, *BRCA2*, *MAP3K3*, *MCPH1*, *RASGRP3*, *PTCH1*, *RDA9B*, *CTNND1*, *ATM*, *NF1*, *KIT*, *PTEN* and *FBXW7*. Many of these mutations were nonsynonymous missense, stop-gain, or frameshift mutations that were damaging. We then performed RNASeq on a subset of these samples to confirm the mutations identified by exome sequencing. This confirmation study verified the presence of a significantly greater

Correspondence to: James V. Tricoli, Ph.D., Cancer Diagnosis Program, Division of Cancer Treatment and Diagnosis, Diagnostic Biomarkers and Technology Branch, National Cancer Institute, 9609 Medical Center Drive, Bethesda, MD 20892 (tricolij@mail.nih.gov).

Disclaimer: Findings and conclusions in this report are those of the authors and do not represent the official position of the National Cancer Institute.

Conflicts of Interest: None for any authors

Authors Contribution Statement: Leadership; (Tricoli), Planning; (Tricoli, Boardman, Furman, Conley, Williams, Khan), Experimental Research Studies; (Tricoli, Lih, Walsh, McGregor, Camalier, Jang), Bioinformatics; (Mehaffey, Patidar, Sindiri), Tissues and Pathology; (Bahrami)

frequency of damaging mutations in AYA compared to adult colon cancers for 5 of these 43 genes, *MYCBP2*, *BRCA2*, *PHLPP1*, *TOPORS* and *ATR*.

Conclusions—Our results provide a rationale for a more comprehensive study with a larger sample set, experimental validation of the functional impact of the identified variants, and their contribution to biological and clinical characteristics of AYA colon cancer.

Keywords

AYA; colon cancer; mutation; exome sequencing; RNASeq

Introduction

One explanation proposed for the relative lack of progress in treatment outcome among the adolescent and young adult (AYA) group of cancer patients is that the biology is different from the corresponding diseases in younger and older individuals not only within the type of cancer, but within the patient. Evidence that colon cancer is more aggressive, and has a poorer prognosis in AYA patients than in older adult patients has existed for some time (1–3). However, the biological basis and the clinical ramifications of this observation remain incompletely defined. Some of the best evidence for a unique biology for AYA colon cancer compared to colon cancer in older adults includes more advanced tumor stage at diagnosis, a greater frequency of mucinous histology and signet ring cells, high microsatellite instability (MSI-H), and mutations of mis-match repair (MMR) genes in AYA colon cancer patients (4–6). In addition, some studies suggest a lower frequency of *KRAS* mutations, reduced frequency of 17p and 18q LOH, and lower p53 protein levels in AYA colon cancer tumors compared to those from adults (7, 8). Several publications within the last two decades support the observation of poorer clinical outcomes for AYA colon cancer patients compared to older adults even when they are placed on similar clinical treatment protocols. In some studies, those patients with non-mucinous tumors displayed a distinct survival advantage over those with tumors displaying a mucinous histology (9). Other studies have shown MSI-H frequency to be greater among AYA colon cancer patients, with a greater percentage of these tumors being mucinous compared to micro-satellite stable (MSS) tumors. One study also revealed no difference between MSI and MSS subgroups with respect to family history of colon cancer, suggesting that MSI may result from acquired rather than inherited genetic defects (10). These data suggest that colon cancer among the AYA age group may be associated with unique clinical and biological properties. A syndrome associated with defects in MMR genes results in young age of onset colon cancer in the absence of polyposis. Lynch syndrome, also sometimes referred to as hereditary non-polyposis colon cancer (HNPCC), is an autosomal dominant syndrome that is associated with an approximately 70% lifetime risk of colon cancer (often right-sided) and a 50–70% percent risk of endometrial cancer (11). The autosomal dominant form is caused by heterozygous mutations in one of four MMR genes, *MSH2*, *MLH1*, *MSH6*, or *PMS2*, and is associated with colon cancer beginning as early as the mid-teens and extending throughout the adult years. A retrospective review of genetic counseling data using a total of 193 patients younger than years of age demonstrated that patients without a hereditary syndrome more

frequently presented with metastatic disease while patients with a syndrome presented at an earlier stage (12).

Two NCI-sponsored workshops have identified AYA colon cancer as a candidate for further analysis to explore a possible unique biology and/or clinical course for this disease compared to that in adults (13, 14). Relatively few molecular genetic studies have been conducted in this age group, perhaps due to the fact that these cases are few in number and tissue samples are difficult to procure. Recent work by The Cancer Genome Atlas (TCGA) study has provided data on genes that are frequently mutated in adult colon cancer (15). This and other studies have identified several genes that exhibit amplification and elevated expression in adult colon cancer, including IGF-2 (16). There are also consensus gene sets that exhibit mutations in adult colon cancer that have been identified (17, 18). These data provide a baseline for pathway analysis that could direct us toward novel signaling pathways in AYA colon cancers. To begin to elucidate any biological distinctions between adult and AYA colon cancer, we have conducted whole exome sequencing on a cohort of 30 adult, 30 AYA, and 2 pediatric colon cancers and detected differences in mutational patterns between AYA and adult samples. We also performed transcriptome analysis using RNASeq on a subset of these samples for which RNA was available in order to confirm the mutations detected by whole exome sequencing, and to investigate expression anomalies.

Methods

Colon Cancer Tissue Samples

The majority of adult and AYA colon cancer tissues were provided by the Mayo Clinic, Rochester MN (LAB). Additional AYA colon cancer tissue samples were provided by St. Jude Children's Research Hospital, Memphis TN (WLF). Samples from adult and AYA colon cancer patients were procured as 10 micron formalin-fixed paraffin-embedded tissue sections with an adjacent hematoxylin-eosin (H&E) stained slide for examination of the tumor pathology. The H&E slides were scanned into an Aperio imaging system for archiving and later analysis. A pathologist examined the H&E stained slides to identify regions that contained 70% or greater tumor tissue content. These regions of tumor were then macro-dissected and placed into bar-coded tubes prior to nucleic acid isolation. Adult colon tumors were from patients aged 61–90 years (mean = 78 years), while AYA colon cancers were from patients aged 19–39 years (mean = 33 years). Two pediatric samples from patients ages 13 and 14 were also analyzed (mean 31 years if you include these two cases). The sample set consisted of 29 males and 33 females. Anatomical sites for the tumors in our sample set include sigmoid, cecum, ascending colon, descending colon, transverse colon, hepatic flexure and splenic flexure are presented in Supplement 1. All tumors were identified as adenocarcinoma with some samples showing evidence of mucinous or signet ring histology. Information on the demographic distribution of the samples is shown in Supplement 2.

Nucleic Acid Isolation and DNA Shearing

DNA and RNA were isolated using the Qiagen Allprep FFPE kit (19). Briefly, the macro-dissected FFPE tissue was de-paraffinized in xylene. The resulting pellet was then digested with proteinase K for 15 minutes with shaking (600 rpm) at 56°C. The solution was

centrifuged and the supernatant transferred to a new tube. The RNA containing supernatant was then digested for a further 15 minutes at 80°C, DNase digested, and purified to completion using Qiagen RNeasy MinElute spin columns. The DNA containing pellet was digested for one hour at 56°C (with shaking) followed by a 2-hour digestion at 90°C (no shaking). The supernatant was treated with RNAse and purified to completion using Qiagen QIAamp MinElute spin columns. After completion, the RNA and DNA concentrations were determined using the Dropsense96 spectrophotometer (Caliper Systems) and stored at −80°C for future use. The quality of the DNA assessed using the KAPA qPCR method (KAPA Biosystems) (20). The results showed a range of values for the Q129/Q41 ratio from 0.110 to 0.652, with the control CEPH cell line having a ratio of 1.01. However, we found that even DNA samples with Q129/Q41 ratios as low as 0.110 yielded libraries of sufficient quality and complexity for whole exome capture and sequencing. In addition, AYA and adult samples were intermixed on the same flow cell and sequencing run, as opposed to sequencing AYA samples on one run and adult samples on another. This was to avoid batch effects from one group compared to the other. DNA was re-quantified using Qubit and sheared on a Covaris at a duty cycle of 10%, an intensity of 5 at 200 cycles per burst for 6 cycles at 60 seconds each at a temperature of 4–7° C., and post-shear cleanup performed using Agencourt beads in preparation for DNA library construction.

DNA Library Construction, Whole Exome Capture and Sequencing

DNA library construction was performed according to the Agilent SureSelect^{XT} Target Enrichment System for Illumina Pair-End Sequencing Library protocol with modifications. Briefly, 500ng input DNA was used for each sample and sheared as described above. Samples were then purified using Agencourt AMPure XP beads and quality assessed using an Agilent 2100 Bioanalyzer system. The ends were repaired, the product bead-purified and A bases added to the 3′ ends of the DNA fragments to generate an A-base overhang. The product of this reaction was again bead purified, and to this indexing-specific paired-end adaptors were attached using T4 ligase. The product of this reaction was bead-purified and the adaptor-ligated library amplified using Herculase II fusion DNA polymerase. The post-amplification product was bead purified and assesses for quality using the Agilent Bioanalyzer system. The library was then hybridized with the SureSelect^{XT} Human All Exon 50Mb library to perform exome capture. Hybridization was performed 65°C for 16 hours and the exome capture library enriched using magnetic Dynabeads (Agilent). Index tags were added by post-hybridization amplification. Exome capture libraries were sequenced on an Illumina HiSeq 2000 using the standard Illumina SOP (21) with modifications.

Exome Sequence Data Analysis

FASTQ files were used to align the sequence to the human genome 19 and mapping, mapping calibration, small variant calling and post-processing performed as described in Supplement 3. The data were filtered for nonsynonymous mutations, mutations displaying greater than 0.01 frequency in the 1000 genomes data base, total coverage of 20 or greater, variant coverage of 4 or greater, GATK quality score of 100 or greater and a variant allele frequency of 20% or greater.

RNASeq Analysis

RNA libraries were prepared according to the manufacturer's instructions for the TruSeq Stranded Total RNA Sample Prep Kit (Illumina, San Diego, CA). The liquid handling Eppendorf (Hamburg, GER) EpMotion 5075 robot was employed for all AMPure bead clean up. Reverse transcription, A-tailing reaction, and adaptor ligation steps were performed manually. Briefly, 1 μg of total RNA was used as input for ribosomal depletion by RiboZero Gold™ (Illumina) to remove both cytoplasmic and mitochondrial rRNA. First strand cDNA synthesis was performed using SuperScript III reverse transcriptase, Actinomycin D, and random primers. Second strand cDNA was synthesized using dUTP. The stranded cDNA ends were A-tailed and ligated with index adaptors for multiplex sequencing. The adapter-modified DNA fragments were enriched by 15 cycles of PCR using primers included in the Illumina Sample Prep Kit. The concentration and size distribution of the libraries were determined on an Agilent Bioanalyzer DNA 1000 chip (Santa Clara, CA). A final quantification, using Qubit fluorometry (Invitrogen, Carlsbad, CA), was done to confirm sample concentration. The libraries were sequenced as 101 paired end reads on an Illumina HiSeq 2000.

Processing of RNA-Seq Data

The paired-end RNA-Seq FASTQ files were subject to quality check by using Prinseq (22). Specifically, the GC content distribution, tag sequence identification, read quality by position, exact duplicates, and 3-end poly A/T distribution for each individual sample was visualized and reviewed manually to set up optimal steps and filtering thresholds for cleaning raw FASTQ files. Based on the manually reviewed results, the following steps were adopted, including remove adaptor sequence, remove read pairs aligned to rRNA or tRNA, remove low average quality reads (Phred score <=30), and trim low quality (Phred score <=30) base pairs from 3-end. The cleaned FASTQ files were analyzed by using RUM (Comparative Analysis of RNA-Seq Alignment Algorithms and the RNA-Seq Unified Mapper RUM. It aligns RNA short reads by mapping with bowtie against both genome and transcriptome. The information is merged and non-mappers are processed by BLAT to do gapped mapping. BLAT and Bowtie mappings are merged for the final alignments. Features are quantified separately for uniquely and non-uniquely mapped reads and junction files are produced.

Differential Expression Analysis

Uniquely mapped reads from RUM workflow were used for differential expression analysis. The expression difference between young and adult onset patients was tested using differential expression analysis for sequence count data (DESeq) (23) that tests for differential expression by use of the negative binomial distribution and a shrinkage estimator for the distribution variance.

Results

Tissue Samples and Quality Metrics

Adult colon tumors from patients ages 61–90 years (median 78 years) and AYA colon cancers from patients ages 19–39 years (median 33 years) were analyzed for mutational differences using WES. Two pediatric samples from patients ages 13 and 14 were also included in the analysis and were counted as AYA samples. DNA isolated from CEPH cell line NA12753 was used as a reference for DNA sequencing. The sample designations, patient age and gender, along with tumor location, histology, tumor grade, and TNM staging information for each of the samples are shown in Supplement 1.

Somatic Mutational Comparison of Adult and AYA Colon Cancer

Complete exome sequencing data was obtained from 30 adult, 30 AYA, and 2 pediatric colon tumor samples using the Agilent SureSelect^{XT} Target Enrichment System for Illumina Pair-End Sequencing Library protocol and sequencing on the Illumina HiSeq 2000 platform. The VCFs were analyzed as described in methods and a comparison was made of somatic mutational profiles between the AYA and adult CC tumors. Since neither germ-line or normal tissue was available from these patients, we selected the top 20 most frequently mutated genes in colon adenocarcinoma as cited in the Catalogue of Somatic Mutations in Cancer (COSMIC) (24). We then compared the mutation frequency for these genes in our adult samples with those in the COSMIC database, and in The Cancer Gene Atlas (TCGA) colorectal cancer database (15). The results are presented in Table 1. The mutational frequencies in our adult sample set for 12 of these genes (*TP53*, *APC*, *ATM*, *PTEN*, *SMAD4*, *PIK3CA*, *RB1*, *KIT*, *NF1, CREBBP*, *TRAPP* and *CARD11*) are very similar to those found in the TCGA database, and mutation frequencies for 5 others (*KRAS, FBXW7, PTCH1, ARID1A* and *KMT2D*) are within 50 percentile points to those observed in TCGA. The remaining 3 genes *(BRAF, CDH1* and *MLH1)* exhibit less correspondence with the mutation frequencies reported in the TCGA data. Performing the same comparison with data from COSMIC there is less consistency with our adult mutational frequencies in that only 8 genes (*TP53*, *PIK3CA, CREBBP, ARID1A, KMT2D*, *CDH1*, *TRRAP* and *MLH1*) closely correspond with the frequencies found in COSMIC, and 4 genes (*KRAS, FBXW7, BRAF and CARD11*) have a mutation frequency within 50 percentile points to that observed in COSMIC. The other 8 genes *APC*, *ATM*, *PTEN*, *SMAD4*, *RB1*, *KIT*, *PTCH1* and *NF1*, display little correlation in mutation frequency between that found in the COSMIC database and our adult samples. Some of the discrepancy between our results and these two databases may be attributable to the fact that the data in TCGA is from over 200 colon cancer samples, and that from COSMIC is derived from many hundreds of samples, while our adult data consists of only 30 samples. Also, COSMIC includes cell line and xenograft sources (included in our analysis), and in some cases tumor samples may be contaminated with germ-line variants. Despite this, in general our adult colon cancer sample set the mutational frequencies for these genes are consistent with those found in COSMIC and TCGA for adult colon cancer tumors.

In our sample sets the statistical significance for differences in gene mutation frequency in AYA compared to adult colon cancer samples was calculated using the Fisher Exact Test

(FET) (25). Initially, there were 43 genes in our study that met the criteria of exhibiting a significantly greater mutational frequency in AYA compared to adult colon cancers ($p \leq 0.05$), and based on gene function or evidence from the literature, to have a potential role in colon cancer. The names and known functions of these 43 genes are shown in Table 2, with the functions derived from the GeneCards database (26). Of these genes 13 had a FET score p-value of <0.01 and thus exhibited the greatest differential in mutation frequency between the AYA and adult groups. These include *ROBO1*, *SIPAL1*, *MYCBP2*, *MED12L*, *BRCA2*, *MSH2*, *COL16A1*, *DNAJC2*, *EPHA3*, *LAMB2*, *MAP3K3*, *MCPH1* and *RASGRP3*. Most of these mutations were nonsynonymous missense, stop-gain, or frameshift mutations that were damaging. For example, *ROBO1* exhibited a mutation in 32% of the AYA colon cancer samples and 0% of the adult samples. Of the ten AYA tumors exhibiting *ROBO1* mutations, seven had at least one damaging mutation as predicted by Sift analysis (27), while two had mutations predicted to be tolerated, and one had a mutation for which the impact on the protein was not available. *ROBO1* has been implicated as a potential tumor suppressor gene in a study of human breast, renal cell and small cell lung cancer in which the promotor hypermethylation inactivation of *ROBO1*was associated with these cancer types (28). Another 13 genes including *BCORL1*, *PTCH1*, *ARHGAP31*, *DEPDC5*, *DMTF1*, *EFS*, *MUC13*, *PGR*, *RAD9B*, *TAXIBP1*, *TONSL*, *PHLPP1* and *TOPORS* had a FET score p-values of 0.01 to <0.02. There were 8 genes with p-values of 0.02 to <0.03 (*SIGLEC10*, *ADAP2*, *GRP133*, *AMPH*, *CTNND1*, *FUS*, *GRPR* and *TLE1*), 7 genes with p-values of 0.03 to <0.04 (*ATM*, *PI3C2G*, *HIPK2*, *NF1*, *ATR*, *KIT* and *PTEN*) and 2 genes with a p-value of between 0.04 and 0.05 (*FBXW7* and *INPP5F*).

Association with AYA Colon Cancer

While all of the genes in Table 2 have the potential to be associated with AYA colon cancer, several are of particular interest based on their more frequent mutation in AYA compared to adult colon cancer, and their known functional role in tumorigenic pathways. Focusing on the genes that display the greatest statistically significant difference in mutation frequency between AYA and adult colon cancer samples ($p< 0.01$), *MYCBP2* encodes a protein that mediates protein ubiquitination and may regulate transcriptional activation of *MYC* (29). This gene displays a mutation frequency of 52% in AYA colon cancer samples and only 7% in adult samples. This compares with a mutation frequency of 7% for this gene in the TCGA database and 8% in the COSMIC database for adenocarcinoma of the colon. *BRCA2* is a tumor-suppressor gene involved in double stranded DNA repair (30) and displays a mutation frequency of 39% in AYA colon cancer and 3% in adult colon cancer compared with mutation frequencies of 5% and 11% in TCGA and COSMIC, respectively. *MSH2* encodes a protein involved in post-replicative DNA mis-match repair (31, 32) and is known to be more frequently mutated in AYA CC and in patients with Lynch Syndrome (12). *MSH2* has a mutation frequency of 39% in our AYA CC sample set, and only 3% in the adults, while the frequencies displayed for TCGA and COSMIC databases were 3% and 8%, respectively. *MAP3K3* encodes a tyrosine kinase that regulates SAPK and ERK pathways by activating SEK and MEK1/2, respectively (33). It was mutated in 26% of the current AYA colon cancer samples, in 0% of the current adult samples, and in only 2% of samples in TCGA and COSMIC data-bases. *MCPH1* encodes microcephalin 1 that arrests cell division at the G2/M checkpoint via CDK1 phosphorylation in response to DNA damage (34). This gene was

mutated in 26% of our current AYA colon cancer samples, in 0% of our current adult samples, and in only 3% and 2% in TCGA and COSMIC databases, respectively. The last gene in this category is *RASGRP3* that encodes the RAS activator RAS guanyl releasing protein 3 (35) and is mutated in 26% of AYA samples and 0% of adult samples. Corresponding mutation rates for this gene in TCGA and COSMIC databases are 3% and 2%, respectively. These results and those for other genes of interest with statistically significant differences in mutational frequency between AYA and adult colon cancer (p= 0.01 to 0.05) that may be involved in colon cancer are reported in Table 3.

There are several other genes known to be involved in colon cancer that exhibit some degree of mutation in our sample set and are shown in Table 4. Of these, *APC* exhibits the greatest frequency of mutation in both AYA and adult samples with 75% of AYA and 73% of adult colon cancer samples having at least one mutation in this gene. Both *MSH2* (36, 37) and *MSH6*, two genes involved in DNA mismatch repair, display large differences in mutation frequency between AYA and adult colon cancer. These genes are known to exhibit an increased mutational frequency in AYA colon cancer. The mutation frequencies in our adult colon cancer samples align well with those found in the TARGET and COSMIC databases in this gene set (Table 4), supporting the validity of our sequencing results despite not having DNA from control tissues to sequence. Since the majority of cases in these databases are likely from adult patients, this provides additional confidence that the increased mutation frequencies we observe in our AYA samples are not sequencing artifacts. The *PTCH1* gene encoding the receptor for the tumor suppressor sonic hedgehog (Tables 2 and 4), also exhibits a statistically significant difference in mutation frequency between AYA and adult colon cancer (38).

Transcriptome (RNASeq) Analysis of Adult and AYA Colon Cancer

The results discussed above are based only on mutation frequency as determined by whole exome sequencing (WES). To confirm these results, we performed transcriptome analysis using RNASeq as described in the methods section. However, only 17 of the AYA colon cancer samples, and 14 of the adult samples had enough high quality RNA on which to perform RNASeq. Of the 43 genes in Table 2 identified by WES as being more frequently mutated in AYA patients compared with adults, 11 were confirmed by RNASeq as being more frequently mutated at the same nucleotide positions as found in the exome sequence data. The results of this analysis are shown in Table 5. The data for these 11 genes was then subject to the additional restrictions of only counting damaging mutations. When this additional restriction was applied only 5 genes maintain a significantly increased frequency of mutation for AYA samples compared with those from adults; *MYCBP2*, *BRCA2*, *PHLPP1*, *TOPORS* and *ATR*. Detailed information on the mutations detected in these five genes and *POLE* are presented in Supplement 4. The genes of interest for AYA colon cancer as defined by our analysis can be parsed by three levels of rigor. Level 1, those genes identified as being more frequently mutated in AYA patients based only on WES and a Fisher Exact Test (FET) score of $p < 0.05$. Level 2, those genes meeting level 1 criteria, and that are validated by RNASeq. Level 3, those genes meeting level 1 and 2 criteria, and containing damaging mutations. For level 3 the cut off was arbitrarily set at an AYA mutational frequency of >20% leading to 5 genes cited. However, a less stringent cut off

would allow the inclusion of several other genes, including *POLE*. There was no discernable correlation between mutation frequency and either tumor location or patient gender for the AYA samples.

Discussion

The incidence of both colon and rectal cancer is increasing in AYA individuals (39, 40), however current colorectal cancer screening guidelines do not provide for early detection screening unless a young person has a family history of young onset colorectal cancer, or a known hereditary genetic risk (41). Recognition of the genetic underpinnings of AYA colon and rectal cancer is crucial for the development of enhanced colorectal polyp and colorectal cancer screening, and for targeted treatments that may differ from those best suited to the AYA colorectal cancer patient. From our study of AYA and adult colon cancer cases there are genes mutated to a significantly higher frequency in AYA colon cancers compare both to our adult sample set and the mutational data reported in TCGA and COSMIC. Based on our analysis of mutations identified by exome sequencing that have been confirmed by RNASeq, the strongest evidence for a greater mutational frequency in AYA colon cancers compared to those from adults is for *MYCBP2*, *BRCA2*, *PHLPP1*, *TOPORS* and *ATR*. This is not to suggest that any of the other candidate genes in Table 2 are not more frequently mutated in AYA, rather that we do not have compelling RNASeq confirmatory evidence for this in the subset of samples for which RNA was available. In this regard, a number of the genes reported in Table 2 are involved in DNA repair pathways. We observe that 39% of AYA samples contained at least one mutation in the DNA mismatch repair gene *MSH2*, while only 3% of adult samples did. While this was expected due to MSI being present in many AYA colon cancers, we also observed a greater frequency of mutations in AYA patients for genes associated with other DNA repair pathways, including *BRCA2* (AYA samples, 39%; adult samples, 3%) and *RAD9B* (AYA samples, 22%; adult samples, 0%), and the cell-cycle checkpoint kinases *ATM* (AYA samples, 35%; adult samples, 7%) and *ATR* (AYA samples, 48%; adult samples, 13%). Therefore, mutations in DNA-repair genes other than the MMR genes *MLH1*, *MSH2* and *MSH6* may be important to the biology and development of colon cancer in the AYA patient population. The WES data (Table 2) shows that *MUC13* is mutated in 23% of the AYA cancers and 0% of the adult and is consistent with the mucinous nature of many AYA colon cancers (42), however altered Muc13 expression was not evident in the RNASeq data.

In a study published by Kothari et al. 2016 a targeted panel of 1321 genes was used to perform exome sequencing on CRC tumors from 195 older (≥ 65 years) and 30 younger (≤45 years) patients. In that study only two genes, *FBXW7* and *POLE*, were found to have significantly greater mutation rates in younger patients (43). However, only 17 of the 43 genes we identified as having a greater mutation frequency in AYA patients were on their 1321 gene panel. Within the aforementioned study's validation set of 21 younger patients, none of the other 26 genes in our panel were identified as more frequently mutated in younger patients. While the Kothari at al. study contained a larger cohort of older patients than ours, the initial set of younger patients (excluding the 21-patient validation set) was the same as ours (N= 30). Of note, the age of the younger patient cohort in the Kothari et. al. study ranged from 30–45 with a median age of 42 (technically outside the upper limit of 39

for AYA patients) while the age of our AYA cohort ranged from 19–39 with a median age of 33. Thus, our younger cohort was more representative of AYA patients. Despite these differences we have also identified *FBXW7* and *POLE* as being more frequently mutated in the younger cohorts than adults. While *FBXW7* was at the edge of statistical significance based on WES data, and was not confirmed by RNASeq, *POLE* is one of the genes in Table 5. While it exhibits a smaller mutational frequency difference between AYA and adult colon cancer, after filtering for validation by RNASeq and for damaging mutations *POLE* achieved the same frequency of mutation in AYA samples as *PHLPP1*, *TOPORS* and *ATR*. Thus, we have included it in Supplement 4 describing the nature of the mutations in the genes of interest. Two *POLE* mutations identified in our AYA samples, P286R and P697A (see Supplement 4 for details), overlap with *POLE* mutations identified in young patients in the Kothari et al. study.

The goal of this study is to provide the research community with an initial and very preliminary landscape of genes that appear to display a greater mutational frequency in AYA colon cancer compared to their adult counterparts. The sample set in our study is small, and derived mostly from a single region of the country. Further scrutiny on larger more comprehensive sample sets may eliminate many of the genes identified here from contention, and prove others not identified here to be more significantly associated with AYA colon cancer. However, the dissemination of these data may eventually reveal gene mutation profiles that demonstrate biological and/or clinical differences between the disease in AYA and adult patients that can then be exploited to better understand AYA colon cancer and better treat these patients. The five genes cited here as exhibiting a significantly greater mutational frequency in AYA colon cancer compared to adult colon cancer may not all be informative in distinguishing the characteristics of AYA colon cancer, or even play a role in colon cancer development. Determining the validity of these findings and deciphering their role in AYA colon cancer development and progression will require larger studies that include an analysis of germ-line sequence and the inclusion of an extensive validation set. Eventually, focused *in vivo* and *in vitro* studies on specific genes will be required to discern the functional role of their products in AYA colon cancer. Only then can we begin to understand the biological differences in AYA colon cancer and apply that knowledge for patient benefit.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We wish to acknowledge the contributions of Dr. James H. Doroshow for funding support, Dr. Xiaoyang Ruan for bioinformatics assistance, Ms. Ruth A. Johnson, Dr. Brooke R. Druliner, Dr. Young Song for pathology support, and Dr. Lisa A. McShane for consultation on the study design.

Funding: This work was supported by internal NCI funding through the Division of Cancer Treatment and Diagnosis.

References

1. Howlader, N.Noone, AM.Krapcho, M.Garshell, J.Neyman, N.Altekruse, SF.Kosary, CL.Yu, M.Ruhl, J.Tatalovich, Z.Cho, H.Mariotto, A.Lewis, DR.Chen, HS.Feuer, EJ., Cronin, KA., editors. SEER Cancer Statistics Review, 1975–2010. National Cancer Institute; Bethesda, MD: Apr. 2013 http://seer.cancer.gov/csr/1975_2010/, based on November 2012 SEER data submission, posted in the SEER web site
2. Bleyer A, Barr R, Hayes-Lattin B, et al. The distinctive biology of cancer in adolescents and young adults. Nat Rev Cancer. 2008; 8:288–298. [PubMed: 18354417]
3. Liang JT, Huang KC, Cheng AL, et al. Clinicopathological and molecular biological features of colorectal cancer in patients less than 40 years of age. Br J Surgery. 2003; 90:205–214.
4. Liu B, Farrington SM, Petersen GM, et al. Genetic instability occurs in the majority of young patients with colorectal cancer. Nature Med. 1995; 1:348–352. [PubMed: 7585065]
5. Kakar S, Aksoy S, Burgart LJ, et al. Mucinous carcinoma of the colon: correlation of loss of mismatch repair enzymes with clinicopathologic features and survival. Modern Path. 2004; 17:696–700.
6. Durno C, Aronson M, Bapat B, Cohen Z, Gallinger S. Family history and molecular features of children, adolescents, and young adults with colorectal carcinoma. Gut. 2005; 54:1146–1150. [PubMed: 15845562]
7. Rajagopalan H, Nowak MA, Vogelstein B, Lengauer C. The significance of unstable chromosomes in colorectal cancer. Nature Rev Cancer. 2003; 3:695–701. [PubMed: 12951588]
8. Popat S, Hubner R, Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis. J Clin Oncology. 2005; 23:609–618.
9. Hill DA, Furman WL, Billups CA, et al. Colorectal carcinoma in childhood and adolescence: a clinicopathologic review. J Clin Oncol. 2007; 25:5808–5814. [PubMed: 18089879]
10. Sultan I, Rodriguez-Galindo C, El-Taani H, et al. Distinct features of colorectal cancer in children and adolescents: a population-based study of 159 cases. Cancer. 2010; 116:758–765. [PubMed: 19957323]
11. Lynch HT, Lynch JF, Lynch PM, Attard T. Hereditary colorectal cancer syndromes: molecular genetics, genetic counseling, diagnosis and management. Fam Cancer. 2008; 7:27–39. [PubMed: 17999161]
12. Mork ME, You YN, Bannon SA, et al. High prevalence of hereditary cancer syndromes in adolescents and young adults with colorectal cancer. J Clin Oncol. 2015; 33:3544–3549. [PubMed: 26195711]
13. Tricoli JV, Seibel NL, Blair DG, Albritton K, Hayes-Lattin B. Unique characteristics of adolescent and young adult acute lymphoblastic leukemia, breast cancer, and colon cancer. J Nat Cancer Inst. 2011; 103:628–635. [PubMed: 21436065]
14. Tricoli JV, Blair DG, Anders CK, et al. Biological and clinical characteristics of adolescent and young adult cancers: acute lymphoblastic leukemia, colorectal cancer, breast cancer, melanoma and sarcoma. Cancer. 2016; 122:1017–1028. [PubMed: 26849082]
15. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012; 487:330–337. [PubMed: 22810696]
16. Tricoli JV, Rall LB, Karakousis CP, et al. Enhanced levels of insulin-like growth factor messenger RNA in human colon carcinomas and liposarcomas. Cancer Res. 1986; 46:6169–6173. [PubMed: 3779638]
17. Sjoblom T, Jones S, Wood LD, et al. The consensus coding sequence of human breast and colorectal cancers. Science. 2006; 314:268–274. [PubMed: 16959974]
18. Wood LD, Parsons DW, Jones S, et al. The genomic landscapes of human breast and colorectal cancers. Science. 2007; 318:1108–1113. [PubMed: 17932254]
19. Qiagen Allprep Kit. Qiagen; Germantown, MD, USA:
20. KapaBiosystems. KAPA Human Genomic DNA Quantification and QC Kit. 2014.
21. Illumina exon capture. Illumina Inc; San Diego, CA, USA:

22. Schmeider R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011; 27:863–864. [PubMed: 21278185]
23. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biology. 2010; 11:R106. [PubMed: 20979621]
24. Catalogue of Somatic Mutations in Cancer (COSMIC). Wellcome Trust Sanger Institute, Genome Research Limited; http://cancer.sanger.ac.uk/cosmic
25. Fisher, RA. Statistical Methods for Research Workers. Oliver and Boyd;
26. Gene Cards. The Weizmann Institute of Science; Copyright 1996–2016. http://www.genecards.org/
27. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Research. 2001; 11:863–874. [PubMed: 11337480]
28. Dallol A, Forgacs E, Martinez A, et al. Tumor specific promoter methylation in the human homologue of the *Drosophila Roundabout* gene *DUTT1* in human cancers. Oncogene. 2002; 21:3020–3028. [PubMed: 12082532]
29. Guo Q, Xie J, Dang CV, Liu ET, Bishop JM. Identification of a large Myc-binding protein that contains RCC-like repeats. Proc Natl Acad Sci. 1998; 95:9172–9177. [PubMed: 9689053]
30. Wooster R, Neuhausen J, Mangion J, Quirk Y, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12–3. Science. 1994; 265:2088–2090. [PubMed: 8091231]
31. Fishel R, Lescoe MK, Rao MRS, Copeland NG. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell. 1993; 75:1027–1038. [PubMed: 8252616]
32. Fishel R, Ewel A, Lee S, Lescoe MK, Griffith J. Binding of mis-match microsatellite DNA sequences to human MSH2 protein. Science. 1994; 266:1403–1405. [PubMed: 7973733]
33. Ellinger-Ziegelbaur H, Brown K, Kelly K, Siebenlist U. Direct activation of the stress-activated protein kinase (SAPK) and extracellular signal-regulated protein kinase (ERK) pathways by an inducible mitogen-activated protein Kinase/ERK kinase kinase 3 (MEKK) derivative. J Biol Chem. 1997; 272:2668–2674. [PubMed: 9006902]
34. Xu X, Lee j, Stearn DF. Microcephalin is a DNA damageresponse protein involved in regulation of CHK1 and BRCA1. J Biol Chem. 2004; 279:34091–34094. [PubMed: 15220350]
35. Aiba Y, Oh-hora M, Kiyonaka S, Kimura Y, et al. Activation of RasGRP3 by phosphorylation of Thr-133 is required for B cell receptor-mediated Ras activation. Proc Natl Acad Sci USA. 2004; 101:16212–16617.
36. Papadopoulus N, Nicolaides N, Wei YF, Ruben SM, et al. Mutation of a mutL homolog in hereditary colon cancer. Science. 1994; 263:1625–1629. [PubMed: 8128251]
37. Aaltonen LA, Peltomaki P, Leach FS, Sistonen P. Clues to the pathogenesis of familial colorectal cancer. Science. 1993; 260:812–816. [PubMed: 8484121]
38. Johnson RL, Rothman AL, Xie J, Goodrich LV, et al. Human homolog of Patched, a candidate gene for the basal cell nevus syndrome. Science. 1996; 272:1668–1671. [PubMed: 8658145]
39. Bailey CE, Hu CY, You YN, et al. Increasing disparities in the age-related incidences of colon and rectal cancers in the United States. JAMA Surg 1975–2010. 2015; 150:17–22.
40. Siegel RL, Fedewa SA, Anderson WF, et al. Colorectal cancer incidence patterns in the United States, 1974–2013. JNCI. 2017; 109:djw322.
41. Bibbins-Domingo K, Grossman DC, Curry SJ, Davidson KW, et al. US Preventive Services Task Force. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. JAMA. 2016; 315:2564–75. Erratum in: JAMA. 2016 Aug 2;316(5):545. PMID: 27304597. DOI: 10.1001/jama.2016.5989 [PubMed: 27304597]
42. Walsh MD, Young JP, Leggett BA, et al. The MUC13 cell surface mucin is highly expressed by human colorectal carcinomas. Hum Pathol. 2007; 38:883–892. [PubMed: 17360025]
43. Kothari N, Teer JK, Abbott AM, Srikumar T, et al. Increased incidence of FBXW7 and POLE proofreading domain mutations in young adult colorectal cancers. Cancer. 2016; 122:2828–2835. [PubMed: 27244218]

Table 1

Percent Mutational Frequency in Colon Cancer Genes

Gene	% Mutated in Our AYA Samples (N=32)	% Mutated in Our Adult Samples (N=30)	% Mutated in COSMIC (N is Variable)	% Mutated in TCGA (N=224)
TP53	61	53	49	55
APC	74	73	42	72
KRAS	35	23	35	42
ATM	35	7	34	12
PTEN	26	3	28	4
SMAD4	29	10	26	12
PIK3CA	45	20	22	15
RB1	16	3	21	4
FBXW7	39	10	20	18
KIT	26	3	20	5
PTCH1	35	3	21	8
NF1	29	3	20	4
CREBBP	29	13	19	10
ARID1A	29	20	17	11
BRAF	10	23	14	9
KMT2D (MLL2)	45	13	16	7
CDH1	16	13	16	3
TRRAP	26	13	15	8
MLH1	10	13	15	3
CARD11	13	7	14	6

Comparison of the top twenty most frequently mutated genes in colon cancer according to the catalogue of somatic mutations in cancer (COSMIC) data base with the mutation frequency of our AYA and adult samples, and TCGA colon cancer samples.

Table 2

Genes with the Greatest Mutational Frequency Difference between AYA and Adult Colon Cancer

Gene	% Mutated in Our AYA Samples (N=32)	% Mutated in Our Adult Samples (N=30)	FET p-value	Gene Name	Gene Function
ROBO1	32	0	0.003763	Roundabout, Axon Guidance Receptor	Neuronal cell migration, deleted in SCLC and breast cancer cell lines
SIPA1L1	32	0	0.003763	Signal-Induced Proliferation-Associated 1	Stimulates the GTPase activity of RAP2A
MYCBP2	52	7	0.00532	MYC Binding Protein 2	E3 ubiquitin-protein ligase, facilitate MYC
MED12L	42	3	0.00533	Mediator Complex Subunit 12-Like	Coactivation of nearly all RNA polymerase II-dependent genes
BRCA2	39	3	0.005683	Breast Cancer 2, Early Onset	Tumor suppressor, ds DNA repair
MSH2	39	3	0.005683	MutS Homolog 2	Post-replicative DNA mismatch repair
COL16A1	26	0	0.008058	Collagen, Type XVI, Alpha 1	Mediates cell attachment and induces integrin-mediated cellular reactions
DNAJC2	26	0	0.008058	DnaJ (Hsp40) Homolog, Subfamily C, Member 2	Molecular chaperone for nascent polypeptides exiting the ribosome
EPHA3	26	0	0.008058	EPH Receptor A3	Ephrin protein-tyrosine kinase, implicated in mediating developmental events
LAMB2	26	0	0.008058	Laminin, Beta 2 (Laminin S)	Cell adhesion, differentiation, migration, signaling, neurite outgrowth and metastasis
MAP3K3	26	0	0.008058	Mitogen-Activated Protein Kinase Kinase Kinase 3	Regulates SAPK and ERK pathways by activating SEK and MEK1/2, respectively
MCPH1	26	0	0.008058	Microcephalin 1	DNA damage response, G2/M checkpoint arrest via CDK1 phosphorylation
RASGRP3	26	0	0.008058	RAS Guanyl Releasing Protein 3	Ras activator
BCORL1	35	3	0.0101	BCL6 Corepressor-Like 1	Transcriptional corepressor, interacts with HDACs
PTCH1	35	3	0.0101	Patched 1	Receptor for sonic hedgehog
ARHGAP31	23	0	0.01512	Rho GTPase Activating Protein 31	GTPase-activating protein, protein trafficking and cell growth
DEPDC5	23	0	0.01512	DEP Domain Containing 5	Component of the GATOR1, inhibits the amino acid-sensing branch of the mTORC1 pathway
DMTF1	23	0	0.01512	Cyclin D Binding Myb-Like Transcription Factor 1	Functions as a tumor suppressor by activating the transcription of ARF
EFS	23	0	0.01512	Embryonal Fyn-Associated Substrate	Possible SH3 domain binding and protein domain specific binding, renal cancer, NSCLC
MUC13	23	0	0.01512	Mucin 13, Cell Surface Associated	Secreted and cell surface glycoprotein associated with colon cancer

Cancer. Author manuscript; available in PMC 2019 March 01.

Gene	% Mutated in Our AYA Samples (N=32)	% Mutated in Our Adult Samples (N=30)	FET p-value	Gene Name	Gene Function
PGR	23	0	0.01512	Progesterone Receptor	Steroid receptor, central role in reproductive events
RAD9B	22	0	0.01512	Checkpoint Clamp Component B	DNA repair exonuclease
TAX1BP1	23	0	0.01512	Tax1 (Human T-Cell Leukemia Virus Type I) Binding Protein	Inhibits TNF-induced apoptosis by mediating the TNFAIP3 anti-apoptotic activity
TONSL	23	0	0.01512	Tonsoku-Like, DNA Repair Protein	Negative regulator of NF-kappa-B mediated transcription
PHLPP1	42	7	0.01778	PH Domain And Leucine Rich Repeat Protein Phosphatase 1	Promotes apoptosis by dephosphorylating and inactivating the serine/threonine kinase Akt
TOPORS	32	3	0.01864	Topoisomerase I Binding, Arginine/Serine-Rich, E3 Ubiquitin	Probable tumor suppressor that regulates p53/TP53 stability
SIGLEC10	55	13	0.02058	Sialic Acid Binding Ig-Like Lectin 10	Adhesion molecule, mediates sialic-acid dependent binding to cells
ADAP2	19	0	0.02925	ArfGAP With Dual PH Domains 2	GTPase-activating protein for the ADP ribosylation factor family, gastrointestinal lymphoma
ADGRD1 (GRP133)	19	0	0.02925	G Protein-Coupled Receptor 133	Transduce extracellular signals through heterotrimeric G proteins
AMPH	19	0	0.02925	Amphiphysin	Associated with the cytoplasmic surface of synaptic vesicles, breast cancer association
CTNND1	19	0	0.02925	Catenin (Cadherin-Associated Protein), Delta 1	Armadillo family protein, functions in adhesion between cells and signal transduction, RCC
FUS	19	0	0.02925	Fused In Sarcoma	Protein component of the heterogeneous nuclear ribonucleoprotein complex, involved in pre-mRNA splicing
GRPR	19	0	0.02925	Gastrin-Releasing Peptide Receptor	A 7-transmembrane G-protein coupled receptor, activates the phopspholipase C pathway, potent mitogen for neoplastic tissues
TLE1	19	0	0.02925	Transducin-Like Enhancer Of Split 1	Transcriptional corepressor, inhibits NF-kappa-B-regulated gene expression, transcriptional activation by FOXA2, and by CTNNB1
ATM	35	7	0.03214	Ataxia Telangiectasia Mutated	Cell cycle checkpoint kinase, phosphorylates CHEK2, p53/TP53, FANCD2, NFKBIA, BRCA1
PIK3C2G	35	7	0.03214	Phosphatidylinositol-4-Phosphate 3-Kinase	Phosphoinositide 3-kinase, cell proliferation, oncogenic transformation, cell survival
HIPK2	29	3	0.03536	Homeodomain Interacting Protein Kinase 2	Serine/threonine kinase function as both a corepressor and a coactivator, phosphorylates p53 PDX1, ATF1, PML and others
NF1	29	3	0.03536	Neurofibromin 1	Negative regulator of the ras signal transduction pathway.
ATR	48	13	0.03596	Ataxia Telangiectasia And Rad3 Related	PI3/PI4-kinase, phosphorylates CHK1, RAD17, RAD9 checkpoint proteins and BRCA1

Gene	% Mutated in Our AYA Samples (N=32)	% Mutated in Our Adult Samples (N=30)	FET p-value	Gene Name	Gene Function
KIT	26	3	0.03763	V-Kit Hardy-Zuckerman 4 Feline Sarcoma Viral Oncogene Homolog	Tyrosine-protein kinase that acts as cell-surface receptor for the cytokine KITLG/SCF, activates AKT signaling pathway
PTEN	26	1	0.03763	Phosphatase And Tensin Homolog	Tumor suppressor, phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase, negatively regulating AKT/PKB signaling pathway
FBXW7	39	10	0.04737	F-Box And WD Repeat Domain Containing 7	Component of a ubiquitin protein ligase complex, phosphorylation-dependent ubiquitination of Myc
INPP5F	39	10	0.04737	Inositol Polyphosphate-5-Phosphatase F	Hydrolyzes phosphatidylinositol 4,5-bisphosphate, modulates AKT/GSK3B

Gene function information derived from Gene Cards, Weizman Institute, (FET) Fisher exact test, $p < 0.050$.

Table 3

Genes Associated with Colon Cancer Exhibiting Significant Mutational Differences in AYA Colon Cancer

Gene	% Mutated in Our AYA Samples (N=32)	% Mutated in Our Adult Samples (N=30)	% Mutated in COSMIC (N is Variable)	% Mutated inTCGA (N=224)	FET p-value	Gene Name
MYCBP2	51.6	6.7	8.5	6.7	0.00532	MYC Binding Protein 2
BRCA2	39	3.3	11.4	5.4	0.005683	Breast Cancer 2, Early Onset
MSH2	39	3.3	7.7	3.1	0.005683	MutS Homolog 2
MAP3K3	26	0	2.3	1.8	0.008058	Mitogen-Activated Protein Kinase Kinase Kinase 3
MCPH1	26	0	2	3	0.008058	Microcephalin 1
RASGRP3	26	0	2.3	3.1	0.008058	RAS Guanyl Releasing Protein 3
PTCH1	35.5	3.3	15.9	7.1	0.0101	Patched 1
PHLPP1	41.9	6.7	2.2	2.7	0.01778	PH Domain And Leucine Rich Repeat Protein Phosphatase 1
SIGLEC10	54.8	13.3	2.3	4.9	0.02058	Sialic Acid Binding Ig-Like Lectin 10
CTNND1	19.4	0	4.4	2.2	0.02925	Catenin (Cadherin-Associated Protein), Delta 1
ATM	35.5	6.7	23.6	11.2	0.03214	Ataxia Telangiectasia Mutated
NF1	29	3.3	14.6	3.6	0.03536	Neurofibromin 1
ATR	48.4	13.3	6.7	4	0.03596	Ataxia Telangiectasia And Rad3 Related
KIT	25.8	3.3	12.3	3.6	0.03763	V-Kit Hardy-Zuckerman 4 Feline Sarcoma Viral Oncogene Homolog
PTEN	25.8	3.3	21	3.1	0.03763	Phosphatase And Tensin Homolog
FBXW7	38.7	10	14.8	12.9	0.04737	F-Box And WD Repeat Domain Containing 7
INPP5F	38.7	10	2.6	2.7	0.04737	Inositol Polyphosphate-5-Phosphatase F

Compares the mutational frequency of those genes we find to be most frequently mutated our AYA colon cancer samples with the mutational frequencies of these same genes in our adult colon cancer samples, and in adult colon cancer samples in the TCGA and COSMIC data bases.

Table 4

Genes Associated with Colon Cancer Genetics

Gene	% Mutated in Our AYA Samples (N=32)	% Mutated in Our Adult Samples (N=30)	% Mutated in COSMIC (N=Variable)	% Mutated in TCGA (N=224)
MLH1	9.7	13.3	9.7	2.7
MSH2	38.7	3.3	7.7	3.1
MSH6	25.8	6.7	10.7	4.5
PMS2	41.9	33.3	2.1	4
EPCAM	12.9	0	<1	0
APC	74.2	73.3	42.4	72.3
PTCH1	35.5	3.3	15.9	8
MUTYH	0	0	1.2	2.7
CTNNB1	12.9	16.7	9.2	4.9

Comparison of mutational frequency data for genes associated with the genetics of colon cancer in the TCGA and COSMIC data bases to the mutational frequency found in our adult and AYA colon cancer samples.

Table 5

Mutational Frequency Data for Genes of Interest Filtered as Described in the Manuscript

Gene	AYA Mutations	Adult Mutations	AYA Mutations Validated by RNASeq	Adult Mutations Validated by RNASeq	AYA Mutations Validated by RNASeq and are Damaging	Adult Mutations Validated by RNASeq and are Damaging
MYCBP2	50% (N=16)	7% (N=2)	47% (N=8)	14% (N=2)	41% (N=7)	0% (N=0)
SIPA1L1	32% (N=10)	0% (N=0)	35% (N=6)	0% (N=0)	12% (N=2)	0% (N=0)
BRCA2	38% (N=12)	3% (N=1)	47% (N=8)	14% (N=2)	29% (N=5)	0% (N=0)
PTCH1	34% (N=11)	3% (N=1)	24% (N=4)	0% (N=0)	18% (N=3)	0% (N=0)
DEPDC5	22% (N=7)	0% (N=0)	29% (N=5)	0% (N=0)	18% (N=3)	0% (N=0)
TAX1BP1	22% (N=7)	0%(N=0)	18% (N=3)	0% (N=0)	12% (N=2)	0% (N=0)
PHLPP1	41% (N=13)	7% (N=2)	24% (N=4)	0% (N=0)	24% (N=4)	0% (N=0)
TOPORS	31% (N=10)	3% (N=1)	29% (N=5)	0% (N=0)	24% (N=4)	0% (N=0)
ATR	48% (N=15)	13% (N=4)	41% (N=7)	7% (N=1)	24% (N=4)	0% (N=0)
POLE	22% (N=9)	13% (N=4)	18% (N=3)	14% (N=2)	24% (N=4)	0% (N=0)
INPP5F	38% (N=12)	3% (N=1)	24% (N=4)	7% (N=1)	18% (N=3)	0% (N=0)

Comparison of mutational frequency differences between our AYA and adult colon cancer samples based on various levels of rigor. N=number of samples in each group.