Supplemental Methods

Gene Based Mutation Screening:

Genotyping was carried out using TaqMan genotyping assay (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. Specifically, 20ng of genomic DNA was used in 5μl total volume of PCR reaction. Primers and probes were custom designed and purchased through Applied Biosystems. Products of 87bp were amplified using Taqman Genotyping Master Mix (Applied Biosystems) combined with the TaqMan assay mix consists of two allele specific probes and a pair of PCR primers (forward primer-CGTCACCTGAGTTTACATATTGAACCT; reverse primer-CCCCAACCTACCATCAAAAATGACT). Samples were run at 95°C for 10 min followed by 40 cycles of 15s at 95°C and 60s at 60°C. The allele detecting probes were VIC labeled ATCCACGTTCCCCC (T36) and FAM labeled ATCCACATTCCCCC (M36). Following PCR amplification, the allele specific signals were measured on the ABI Prism 7900 using Sequence Detection Software 2.3 (Applied Biosystems). The quality value is calculated using Applied Biosystems' proprietary calling algorithm. Genotyping Quality value was at least 95%. Internal controls were sequence verified WT, T36M and M36 genomic DNAs.

Human Subjects:

IRB approval was obtained for the analysis performed in this study (IRB# 354-06; 356-06, 622-00, 806-96 and 08-007049 for control DNA samples and IRB# 622-00; 332-94 and 806- 96 for the genomic DNAs from subjects with colorectal cancer).

ARPKD patients:

ARPKD patients of European origin were collected mainly from the US and Spain, as previously defined. Patients were included only if they had mutation proven ARPKD or a typical ARPKD phenotype(Rossetti et al. 2003).

Cases:

Colon cancer cases were selected from three prospective collections. The first, the Biobank for Gastrointestinal Health Research, is an ongoing IRB approved collection of biospecimens from participants with normal colonoscopic examinations, colon polyps or colon and rectal cancer seen at Mayo Clinic Rochester between the years of 1995 to 1998 and 2000 to the present (mean age = 63.2; 58% males) . The second group was derived from a subset of the North Central Cancer Treatment Group (NCCTG) clinical trial NO147. Cases for the clinical trial were enrolled at over 65 NCCTG sites in the US and Canada, which included Mayo Clinic (Rochester, Arizona, and Florida). As part of the clinical trial protocol, biospecimens were collected beginning in March, 2004. Patients aged ≥18 years with histologically confirmed and surgically removed stage III colon cancer were eligible. Of the consented patients, only US resident Europeans were included in this study (mean age = 64.4; 55% males). Proximal colon tumors included those arising in the cecum, the ascending, and transverse colon (including both flexures). Distal tumors were defined as those occurring in the descending or sigmoid colon. The third group of cases utilized originated from the Colon Cancer Family Registry, an international consortium of six centers in North America and Australia collaborative resource described in more detail by Newcomb et al., which collected both CRC cases from family clinics, population based sampling from incident cases of CRC and population based recruitment of patients with younger age of cancer onset while non-blood relative and/or spousal controls were collected from these centers (mean age = 49.9; 51% males).

Controls:

Clinic-based controls were collected from the Biobank for Gastrointestinal Health Research via the Clinical core of the Mayo Clinic Center for Cell Signaling in Gastroenterology (P30DK084567), the Colon Cancer Family Registry; the Mayo Clinic SPORE in Pancreatic Cancer (P50 CA102701); and the Mayo Clinic Biobank. Controls of European descent from the Biobank for Gastrointestinal Health Research (consented subjects with normal colonoscopies

and no prior polyp or cancer history) (mean age = 64.2; 56% males), and the Colon Cancer Family Registry are described above in more detail (mean age =50.7; 49% males). The control group provided by the prospective collection of the Mayo Clinic SPORE in Pancreatic Cancer was derived from consented subjects who had been seen for a routine annual general physical examination in the Department of Medicine, Mayo Clinic, Rochester, MN, since 2000-2008 (mean age = 61.2; 46% males) (McWilliams et al. 2005). Initiated in 2009, the Mayo Clinic Biobank controls are subjects from Olmsted County found to be healthy during a medical examination and in the Department of Medicine divisions of Community Internal Medicine, Family Medicine and General Internal Medicine (mean age = 50.4; 51% males). For all control subjects, none had a history of cancer at the time of enrollment and all were of European descent.

Overall, there were 1483 cases ascertained from the two sources described in the previous section: Biobank for Gastrointestinal Health Research (N=834) and NCCTG_N0147 (N=649). One hundred and ten individuals of non-Caucasian ancestry, 60 patients not from the continental US, and 6 others not meeting other inclusion criteria were excluded from this study. Of the 1307 available cases, 1244 were matched on age, gender and state of residence to the general medicine controls (all Caucasian). DNA was not available for 129 cases and 235 controls. This yielded 1178 cases, to which another 693 sporadic colorectal cancer genomic DNA samples were added. Genotyping was performed on a total of 1871 cases (29 failures to amplify 1.6%) and 1621 controls (20 failures to amplify 1.2%).

Statistics: Confidence levels were computed using the central confidence interval calculator at (http://www.causascientia.org/math_stat/ProportionCI.html), and the Fisher's exact test using the calculator at, http://www.langsrud.com/fisher.htm or the R statistics package http://www.r-project.org/