**Supplementary Methods (Ochs-Balcom et al)**

**METHODS**

Utah Population Data Base (UPDB). The UPDB includes Utah genealogy from the Northern European founders in the mid-1800s to modern day1Approximately 3 million of the individuals in UPDB are part of at least 3 generations and connect to a Utah founder.

The Utah Cancer Registry (UCR), established in 1966, became an NCI Surveillance, Epidemiology, and End-Results (SEER) Registry in 1973. Approximately 166,000 cancer records up to 2014 are linked to individuals with at least three generations of genealogy and analyzed.

Colorectal cancer (CRC) cases (n=17,024) with linked genealogy data were identified from SEER codes as cecum (21041), ascending colon (21043), hepatic flexure (21044), transverse colon (21045), splenic flexure (21046), descending colon (21047), sigmoid colon (21048), large intestine NOS (21049), rectosigmoid junction (21051), and rectum (21052). Left-sided CRC cases (n=11,355) included SEER cancer codes for splenic flexure, descending colon, sigmoid colon, rectosigmoid junction and rectum; right-sided CRC cases (n=6,914) included SEER codes for cecum, ascending colon, hepatic flexure and transverse colon. CRC cases coded as 21049: large intestine NOS were included in all CRC cases, but not included in either left- or right-sided cases due to the non-specific localization.

Estimation of Relative Risks in Relatives. All individuals in the UPDB with genealogy data as described were assigned to a sex-, 5-year birth year range, and birth state-(Utah or not) specific cohort. Cohort-specific rates for each phenotype were estimated by counting all of the CRC cases in each cohort and dividing by the total number of individuals in each cohort.. The “N” is the total number of relatives of the 1,510 early-onset CRC probands for each degree of relative, ‘observed’ is the observed number of relatives with early-onset CRC, ‘expected’ is the expected number of relatives diagnosed with early-onset CRC. The significance and 95% CI for the 2-tailed test used the method of Agresti2.

Genealogical Index of Familiality (GIF) Test for Excess Relatedness of Cases. The GIF test allows for testing the hypothesis of excess relatedness among a group of individuals with a common phenotype. The test compares average pairwise relatedness of cases (e.g. CRC dx <50 years) with the expected pairwise relatedness for a matched set of individuals in the UPDB. Coefficient of kinship is used to measure relatedness; pairs are defined based on genetic distance. The expected pairwise relatedness for a set of cases is estimated in a set of randomly selected, matched controls from UPDB. Controls were matched on the cohorts described above. For the GIF test for a set of cases, the expected pairwise relatedness was estimated as the average pairwise relatedness computed for 1,000 sets of matched controls. The significance of the case GIF was empirically assessed by comparison with the 1,000 control GIF values. The GIF method tests excess relatedness, but does not distinguish between relatedness due to genetics versus that due to common environment. For this reason, the distant GIF (dGIF) test is also performed while ignoring all relationships closer than third-degree, allowing a test of excess distant relatedness which is unlikely in the absence of an inherited contribution to the phenotype examined.

**REFERENCES**

1. Cannon-Albright LA, et al. Identification and study of Utah pseudo-isolate populations-prospects for gene identification Am J Med Genet A 2005;137A:269-75.

2. Agresti A. Categorical Data Analysis: Wiley, 1990.