

### Fight CRC EAO Data Repository List

Organization/Name	PI/ Contact Name	Description
<a href="#">Atherosclerosis Risk in Communities Cohort – Cancer (ARIC-Ca)</a>	Elizabeth A Platts: Johns Hopkins Bloomberg School of Public Health	ARIC is a cohort study funded by the National Heart, Lung, and Blood Institute (NHLBI) of 15,792 participants that was designed to investigate the etiology and natural history of atherosclerosis. The cancer component, ARIC-Ca, is sponsored by the NCI to leverage ARIC's wealth of data, the racial diversity of the cohort (27% of participants are African-American), and the cohort's long-term follow-up to enhance cancer epidemiology research, including on cancer incidence, mortality, recurrence, progression, and case-fatality.
<a href="#">Memorial Sloan Kettering Cancer Center</a>	Andrea Cercek	To better understand the differences in etiology and biology of these EO tumors, we retrospectively reviewed the charts of all pts with EO CRC (n = 384) treated at Memorial Sloan Kettering Cancer Center between 2014 and 2017 for pathology, radiology, and clinical characteristics and compared these to a cohort of AO CRC cases (n = 543). This is the largest cohort describing a broad range of clinical and genomic characteristics of the EO CRC population with comparison to AO CRC
<a href="#">Millennium Cohort Study</a>	Rudolph Bull: Department of Defense (DoD) Center for Deployment Health Research, Naval Health Research Center	This is the first large, population-based prospective study to investigate the short- and long-term health effects of military service. After the 1991 Gulf War, research on illnesses was hampered by unavailability of objective measurements on exposures at the individual level, a lack of baseline health data, and an inability to adequately control for potential confounding factors. Public health experts recommended that the Department of Defense (DoD) establish a cohort with the capability to prospectively examine health outcomes among U.S. military service members. The Millennium Cohort Study was developed to fulfill this need.
Multi-institutional cohort in Europe	Richard Boland	Performed testing for MSI/IHC on CRCs in individuals <50
<a href="#">Ohio population-based cohort of CRC patients</a>	Heather Hampel	Population based cohort of CRC patients in the state of Ohio which includes 750 patients diagnosed <50 who are fully molecularly characterized and all had germline genetic testing.
Single institution cohort in Dallas, TX	Richard Boland	Performed testing for MSI/IHC on CRCs in individuals <50
<a href="#">ACCENT Collaborative Group</a>	Quian Shi (Richard Goldberg)	To address important scientific questions in early-stage colon cancer, the Adjuvant Colon Cancer End Points (ACCENT) Collaborative Group has continuously integrated individual patient data from large adjuvant colon cancer RCTs conducted worldwide. Through June 2016, IPD from over 40,000 patients from 27 major trials (1977 to 2009) have been assembled.

<a href="#">ARCAD (Aide et Recherche en Cancérologie Digestive) Clinical Trials Program</a>	Aimery de Gramont	Worldwide collaboration of clinicians, statisticians and scientists specializing in gastrointestinal malignancies. The program's objective is to facilitate the development of more efficient clinical trials in gastrointestinal cancers through consensus recommendations and original research.
<a href="#">Cancer Care Outcomes Research &amp; Surveillance Consortium (CanCORS)</a>	NCI (12 PI's listed in last link)	Supported prospective research in a cohort of approximately 10,000 patients with newly diagnosed lung cancer or colorectal cancer recruited from geographically diverse populations and health care systems.
<a href="#">Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO)</a>	Ulrike Peters	Collaborative effort of researchers from North America, Australia, and Europe, using data from over 40,000 participants. Based at the Fred Hutchinson Cancer Research Center. GECCO aims to accelerate the discovery of colorectal cancer-related variants by replicating and characterizing Genome Wide Association Study (GWAS) findings, conducting a large-scale meta-analysis of existing and newly generated GWAS data, and investigating how genetic variants are modified by environmental risk factors.
<a href="#">International Survival in Colorectal Cancer (ISACC) Consortium</a>	Polly Newcomb, Andy Chan	Includes 23 studies and nearly 27,000 colorectal cancers, in order to initiate research projects with Consortium investigators and mentees to leverage the diverse and well-annotated populations and biospecimens collected by the participating studies.
<a href="#">PROSPR (Population-based Research to Optimize the Screening Process) Consortium</a>	NCI, PI's: PROSPR Coordinating Center PI: Christopher LI, Colorectal PROSPR Research Center, PI: Doug Corley	Population-based Research to Optimize the Screening Process (PROSPR) is a network that is conducting research to better understand how to improve the cancer screening process (recruitment, screening, diagnosis, referral for treatment) in community healthcare settings. UT Southwestern is part of the PROSPR Consortium (along with Kaiser Northern California and Kaiser Washington) and following a cohort of more than 1 million primary care patients through the colorectal cancer screening process, we are expanding our age criteria to 40 so will have access to patients in this age group.
<a href="#">NCI Cohort Consortium</a>		The NCI Cohort Consortium membership is international in scope. It includes investigators responsible for more than 40 high-quality cohorts who are studying large and diverse populations in more than 15 different countries. Multiethnic Cohort, Nurses' Health Study, Health Professionals Study, California Teachers Study, Millennium Cohort Study, Women's Health Initiative, Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, Colon Cancer Family Registry Cohort (CCFRC).
<a href="#">Federal Statistical Research Data Centers</a>	United States Census	There are currently 29 open Federal Statistical Research Data Center (RDC) locations. The RDCs partner with over 50 research organizations including universities, non-profit research institutions, and government agencies.
<a href="#">Health Care Systems Research Network (HCSRN) Virtual Data Warehouse</a>	Steve Waring	The HCSRN's Virtual Data Warehouse (VDW) facilitates multi-site research while protecting patient privacy and proprietary health practice information. Originally developed by the Cancer Research

		Network, the VDW now supports studies of cancer, drug safety, cardiovascular disease, mental health, and more.
<a href="#">National Center for Health Statistics</a>	CDC	The National Center for Health Statistics (NCHS) is a rich source of data for researchers, teachers, and students who want to perform data analysis. Conducting two analyses of ages <50 in the NHCR -- exam outcome comparison and factors associated with advanced adenoma/cancer.
<a href="#">North American Association of Central Cancer Registries (NAACCR) data</a>	Executive Director: Betsy Kohler	population coverage is much higher – approximately 95% for the most current five data years. It is combined incidence from central cancer registries supported by SEER and/or CDC's NPCR, and also includes Canadian registries.
<a href="#">SEER Data</a>	NCI (Becky Siegel)	The Surveillance, Epidemiology, and End Results (SEER) Program provides information on cancer statistics in an effort to reduce the cancer burden among the U.S. population. SEER is supported by the Surveillance Research Program (SRP) in NCI's Division of Cancer Control and Population Sciences (DCCPS).
<a href="#">Veterans Affairs Corporate Data Warehouse</a>	Department of Veterans Affairs (VHA)- Veterans Health Administration	Consolidating data from disparate sources into a coherent single logical data model. The Corporate Data Warehouse (CDW) is the physical implementation of this logical data model at the enterprise level for VHA.
<a href="#">100,000 Genomes Project</a>		~1100 tumours and matched normal samples have been whole genome sequenced. 93 of these are from EO CRC.
<a href="#">UK Biobank</a>	Rory Collins	SNP genotype data on 500,000 participants from the UK. 508 cases of CRC diagnosed before age 50
<a href="#">Spanish hospitals network</a>	Jose Perea Garcia	Collecting, mainly prospectively, EO CRC samples (blood, healthy and tumoral tissues, and stool) and clinical data. Moreover, we are going to give them a questionnaire about lifestyles, habits, personal and familial disease history, diet, etc., in order to identify risk factors related to the EO CRC. Main hospitals in Milan, UK, Luxembourg, Poland, France, and trying to incorporate some more. At this moment 12, 7 of them in Madrid, and 5 more around Spain.
<a href="#">The Nurses' Health Study II</a>	A. Heather Eliassen, Walter Willet	US female nurses aged 25 to 42 years at study enrollment (1989). Among the 85 256 women studied, 114 cases of early-onset CRC were documented
<a href="#">SHINE (Stroke Hyperglycemia Insulin Network Effort)</a>	NIH	SHINE is a multicenter, randomized, controlled clinical trial of 1400 participants that will include approximately 60 sites across the country. The trial will evaluate if blood sugar control is safe and effective for improving stroke recovery. The research sites will be testing the current standard treatment against the administration of a controlled IV insulin infusion therapy in eligible participants who present within 12 hours of the onset. The treatment lasts for up to three days.
<a href="#">CDC Comparative Effectiveness Research (CER) data/ Louisiana Tumor registry</a>	CDC	All CRC patients aged ≤50 years, diagnosed in Louisiana in 2011, who underwent surgery in 2011-2012 were retrospectively analyzed. Prevalence of, and the factors associated with operation type (ECR including subtotal/total/proctocolectomy vs. segmental

		resection) were evaluated. Results: Of 2,427 CRC patients, 274 were aged ≤50 years.
<a href="#">Colon Cancer Family Registry Cohort (CCFRC)</a>	NCI Cohort Consortium (PI's: Mark Jenkins: The University of Melbourns, Jane Figueiredo: Cedars-Sinai Medical Center, Steven Gallinger: Mount Sinai Hospital, Loic Le Marchand: University of Hawaii at Manoa, Noralane M. Lindor: Mayo Clinic, Arizona, Polly A. Newcomb: Fred Hutchinson Cancer Research Ctr)	International consortium of six institutes in the United States, Canada and Australasia formed as a resource to support studies on the etiology, prevention, and clinical management of colorectal cancer. The resource comprises data and biospecimens from over 42,000 participants from more than 15,000 families that have been recruited from 1998 to 2011.
<a href="#">Scripps Green Hospital</a>	Richard Boland	109 EAO CRC individuals included.
<a href="#">Cancer Genetics Network</a>	NCI (14 clinical research centers)	The Cancer Genetics Network was a national research network sponsored by the National Cancer Institute to support collaborative investigations into the genetic basis of cancer susceptibility, explore mechanisms to integrate this knowledge into medical practice, and to identify ways of addressing the associated psychosocial issues. It was formed in 1999, and was funded through 2012. It was comprised of 14 clinical research centers, a data coordinating center, and an informatics group.
<a href="#">ACS Cancer Prevention Study (CPS-3)</a>	Alpa Patel	In December 2013, the American Cancer Society completed the initial recruitment of its newest study, Cancer Prevention Study-3, with over 304,000 participants. The American Cancer Society's Epidemiology Research Program invited men and women between the ages of 30 and 65 years who had no personal history of cancer to join this historic research study. The ultimate goal was to enroll at least 300,000 adults from various racial and ethnic backgrounds from across the United States and Puerto Rico. Part of NCI Cohort Consortium
<a href="#">Colorectal Transdisciplinary (CORECT) Study</a>	Dr. Peters from FHCRC, and Dr. Gruber, Dr. Casey, and Dr. Thomas from University of Southern California.	The CORECT project is investigating and identifying susceptibility loci for colorectal cancer in order to characterize the biologic basis of inherited susceptibility and recognize how genetic variation may be quantified and modified by environmental risk factors. Complex risk models that take advantage of this new genetic information, together with known environmental factors for colorectal cancer, will expedite the translation of these findings into clinical practice.
<a href="#">CORGI (COloRectal Gene Identification)</a>	Ian Tomlinson	In this study, the researchers will look at the genes of a large number of people who have had a tumour in their bowel. They will look for common gene faults that may increase the risk. And they will find out more about people's family history to see if any of their relatives have also had tumours in the bowel. The aim of the study

		is to identify genes that may cause bowel tumors. Currently researchers have whole genome sequencing for 176 EOCRC cases and SNP genotyping arrays for 1500 EOCRC cases.
<a href="#">Patterns of Care (POC) Studies</a>	NCI	Evaluate the dissemination of state-of-the-art cancer therapy and diagnostics into community oncology practice, identify patient-, provider-, and system-level factors that are associated with receipt and utilization of state-of-the-art cancer therapy and diagnostics in community oncology practice; and Disseminate findings through scientific publications and presentations.
<a href="#">NCI Colon and Rectal TF</a>	NCI	The NCI Gastrointestinal Cancer Steering Committee (GISC) was established in January 2006. At monthly meetings, the GISC addresses the design, prioritization, and evaluation of concepts for phase 2 and phase 3 clinical trials in adult gastrointestinal cancers.
<a href="#">Oncology Research Information Exchange Network</a>	M2Gen in Tampa, FL - Bill Dalton	19 cancer centers in US are participating and enrolling cancer patients since 2014. DNA, plasma, serum, fresh frozen tumors, FFPE tumor and normal available along with anything in the EMR. A subset are sent to Hudson Alpha for tumor and normal exome sequencing and tumor RNAseq.
<a href="#">National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)</a>	Director: Griffin P. Rodgers, M.D., M.A.C.P.	Several studies have been funded to explore dietary factors in early childhood; explore relevance to EAO CRC projects. Healthy Start Study, National Children's Health Study
<a href="#">Vanderbilt childhood cancer survival study</a>	Vanderbilt- Ingram Cancer Study researchers	Retrospective cohort study: A large, retrospective study of the children of childhood cancer survivors who were treated with radiation therapy and/or some forms of chemotherapy found that the offspring do not have an increased risk for birth defects compared with children of cancer survivors who did not receive these treatments.
<a href="#">National Longitudinal Survey of Youth of 1997 and 1979</a>	Bureau of Labor Statistics	Survey of young men and women born in the years 1980-84; respondents were ages 12-17 when first interviewed in 1997. Survey of men and women born in the years 1957-64; respondents were 14-22 when first interviewed 1979.
<a href="#">New Hampshire Colonoscopy registry (Dartmouth)</a>	Lynn Butterly	Prospective registry: The primary goal of the study is to learn about how medical, environmental, and inherited factors affect the incidence of colon cancer in New Hampshire.
The University of South Carolina Palmetto Health Biorepository	Phillip Buckhaults, Ph.D. and Charlie Butler, MD.	250+ frozen colon cancers and matched normal. 50 colon cancer and normal organoid cultures. clinical data with outcome.
<a href="#">Exploring Potential Outcomes among Children (EPOCH)</a>	Colorado School of Public Health: Dana Dabelea, Katherine Sauder, Allison Shapiro	Longitudinal study of adolescents and their long-term health outcomes related to whether they were exposed to mother's diabetes during pregnancy. While it is known that a mother with diabetes during pregnancy increases the risk of childhood obesity and diabetes, many other things are not known, including whether infant or childhood feeding changes these outcomes.
<a href="#">KP Multiphasic Evaluation Study</a>	Kaiser Permanente	

<a href="#">The Health Improvement Network (THIN)</a>	UK	THIN database is a large UK primary care database. The research team uses this database containing data collected from over 550 general practitioners spread over the UK for research into cardiovascular disease, mental health, pharmacoepidemiology and other fields of primary care research. Medical records of children born at 1995-2013, enrolled after birth and collect height/weight measures as well as medication records
<a href="#">IBM MarketScan</a>	IBM	They provide researchers with a patient-centric perspective reflecting real-world treatment patterns and the full cost of care. Full integration, longitudinal strength, deep cross-sectional detail, and the excellent documentation of the IBM MarketScan claims databases ensure unparalleled research data quality. All data are fully HIPAA compliant.
Thousand Community Cohort	May be part of the NCI Cohort Consortium	
<a href="#">National Health and Nutrition Examination Survey (NHANES)</a>	CDC, Division Director: Kathryn Porter	The National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations. Access previous exposure data and link to prospective collection of outcomes data.
<a href="#">GI Quality Improvement Consortium Ltd (GIQuIC)</a>	Glenn Eisen (Director and President)	Systematically collected endoscopy data from hospitals, universities, ambulatory surgery centers, and office-based endoscopy units.
<a href="#">All of Us Study</a>	NIH, Director: Eric Dishman	<i>All of Us</i> is a participant-engaged, data-driven enterprise supporting research at the intersection of human biology, behavior, genetics, environment, data science, computation and much more to produce new knowledge with the goal of developing more effective ways to treat disease.
<a href="#">Growing Up Today Study (GUTS)</a>	Rulla Tamimi	GUTS participants are a unique cohort of over 27,000 young adults, for whom there is 20 years of data collected during adolescence and young adulthood in addition to nearly 30 years of data from their mothers, who continue to participate in the Nurses' Health Study II
<a href="#">Nurses Health Study I</a>	Walter Willet	The Nurses' Health Studies are among the largest prospective investigations into the risk factors for major chronic diseases in women.
<a href="#">Pooling Project of Prospective Studies of Diet and Cancer</a>	Stephanie Smith Warner	The Pooling Project of Prospective Studies of Diet and Cancer (DCPP) is an international consortium of 39 cohort studies (see Participating Cohort Studies) established in 1991 to investigate associations between dietary and anthropometric factors and risk of cancer by analyzing harmonized participant-level data from multiple cohorts using standardized criteria across studies.
<a href="#">NCI Community Oncology Research Program (NCORP): Early Onset Malignancies Initiative</a>	Worta McCatskill-Stevens	Aims to investigate why certain racial and ethnic populations are at increased risk of developing cancer at an early age, through the collection and characterization of malignancies.



<a href="#">SEER Residual Tissue Repository (RTR)</a>	W. Atkin, C. Lynch, L. LeMarchand	CRC cases collected at LA, Hawaii, and IA SEER registry sites (no longer funded, but tissues may be available)
<a href="#">Cancer Prevention Study II (CPS-2) Nutrition survey; bio repository</a>	Susan Gapstur	CPS II provides the relative risk estimates from smoking used by the Centers for Disease Control and Prevention (CDC) to estimate deaths attributable to smoking in the United States. More recently, CPS II publications have contributed to our understanding of how various medications, medical conditions, familial, and environmental factors may affect cancer risk.
<a href="#">Childhood obesity intervention study</a>	Harvard T.H. Chan	The study aims to identify: 1) whether the school-based intervention will be effective for preventing excessive weight gain among children; 2) whether the intervention will be beneficial for improving healthy eating, physical activity and reducing sedentary behaviors among children.
<a href="#">ECHO (Environmental Influences on Child Health Outcomes)</a>	NIH, Director: Matthew Gillman	ECHO will support multiple, synergistic, longitudinal studies using existing study populations, called cohorts, to investigate environmental exposures — including physical, chemical, biological, social, behavioral, natural and built environments — on child health and development.
<a href="#">UK Cancer Registry</a>		Cancer cases identified from 1971-2013; Age-period-cohort models; ~ 120k cases
<a href="#">Early Detection Research Network (EDRN)</a>	Robert Schoen (Lead investigator of colorectal-specific project); Ziding Feng (Lead investigator of data management and coordinating center)	Consortium of multiple institutions to accelerate translation of biomarker information into clinical applications and to evaluate new ways of testing cancer in its earliest stages and for cancer risk
<a href="#">Colon Molecular Atlas Project (Colon MAP)</a>	Robert Coffey	Recently awarded Cancer Moonshot grant to map out routes that benign polyps take to progress to cancer; data will be shared with all researchers as part of the Human Tumor Atlas Network (HTAN)
Registry and medical record data from Scandinavian countries	Various	Medical record data linked with disease registries, tracking all encounters and diagnoses among persons with national health insurance in these countries
<a href="#">European COST CA 17118</a>	Sergi CASTELLVI-BEL (Chair)	Identifies biomarkers through translational research for the prevention and stratification of colorectal cancer
<a href="#">PEW Research</a>	Michael Dimock	Pew Research Center is a nonpartisan fact tank that informs the public about the issues, attitudes and trends shaping the world. We conduct public opinion polling, demographic research, content analysis and other data-driven social science research. We do not take policy positions.